

Beyond this winter's solstice

Those who would like to think of this new year as a clean slate had better remember last year's unsolved problems, many of them financial. But there are also new opportunities.

THE convention, followed in several pages in this issue, that periodicals of all kinds take the turning of a year as an occasion for reviewing past events and predicting some that lie ahead is, naturally, pure convention. The job, if worth doing, could be done at any time. In Western European cultures, the near-coincidence with the winter solstice, to which mystical significance used to be attached, may be part of the explanation, but is hardly appropriate in the Southern Hemisphere. It is not even, these days, that there is nothing else to write about. Nor, unfortunately, is it any longer possible to pretend that the past can be put aside and that the future will be free-standing and self-contained. The instability of the international financial markets will, for example, continue in the year ahead, threatening to undermine people's prosperity, the capacity of governments and companies to support research and, more generally, international amity, even civility.

With that gloomy prospect, it is natural to fall back on a listing of hopeful trends. The most striking of these is represented by the signature last November of the treaty between the United States and the Soviet Union that will put an end to the deployment of nuclear missiles of intermediate range. The two parties now say they hope to complete an agreement on strategic missiles in the first half of this year. The plan is to reduce the deployment of nuclear warheads carried by long-range missiles by roughly a half. That, if it materializes, will be a great achievement. But even an agreement on strategic missiles will not substantially bring closer the unattainable ideal of a nuclear-free world; even half of present strategic nuclear forces will be more than enough to effect the promise of unacceptable damage required by the classical theories of strategic deterrence, there are in any case other nuclear powers and the prospect of non-nuclear warfare grows steadily more awesome (and costly).

The prospect of further agreements on nuclear weapons between the superpowers is cheering because, if they can be made safe and dependable, there will then be resources to spend on more useful things and there is a high chance that arms control agreements may pave the way for more general understanding. What the research community has most to gain is the much fuller intercourse between its two halves, in the East and the West. The value of that prize is incalculable but great.

Research

With some local exceptions, the environment for research is, for the time being, healthy. (The effect of continuing pressure on the US dollar on federal support for research will need watching.) Most governments have slipped into believing that the future of their societies rests on their capacity to compete for markets for technological products that have not yet been invented. The United States over the Reagan years has been the most consistent in its support for imaginative research, but is still at odds with itself on government support for applied research (which, on paper, is disallowed, but which is a by-product of federal programmes in fields such as defence, energy and space). Might it not be a worthwhile resolution for the new year, even by a government at the end of its lease, that it should moderate some of its ambitions, the plan for building a space

station for example? Much the same, but more so, goes for Europe. It is beyond belief that European governments, which have many other things to do, should have fallen in with a plan to send Europeans into space which is itself a relic of now-faded French enthusiasm for all things technological (see p. 8).

The Soviet Union's problem is quite different. The paradox that a rich country able also to claim superpower status on the strength of technological developments in the military field should be economically crippled by the weakness of its civilian industry (and agriculture) is most simply explained by the absence of a market. Mr Mikhail Gorbachev and his associates are seeking a substitute for such a means of ensuring that economic resources are indeed channelled in efficient directions. (The suggestion last weekend that Soviet workers may be encouraged to own shares in the enterprises for which they work is part of that process.) The difficulty, as the Soviet leadership well knows, is that ultimately there may be no substitute for a market, which would be politically embarrassing. Luckily, there is much that may be done more quickly by structural and administrative reform. The benefits, for the rest of us as well as the Soviet government, of liberating an army of talented researchers from the shackles of a deadening bureaucracy would be immense.

Parochial problems unfortunately obtrude. Australia's celebration of its bicentenary also marks the ex-British dominion's adoption of modern British precepts on the support of research, and especially the conviction that basic research mostly makes sense only in partnership with industry. The difficulty, perhaps historical and cultural in origin, is that both countries are blessed with putative industrial partners whose conviction that research is beneficial is moderated by their unwillingness to spend real money on it, or to employ technical people intelligently. The climate in Britain is changing slowly, but it will take some time to persuade Australians to be other than primary producers.

The outstanding British problem for 1988 is not merely a continuation of past difficulties, but a consequence of the misnamed Education Reform Bill, published at the end of last year. After more than eight years in office, the British government can justly point to deliberately engineered economic growth during a period when the volume of original research has been, if anything, shrinking. There is a chance (spending estimates due in February will confirm this) that the long period of financial harassment is about to end. All that is welcome. But the government's proposals on education, which would give this and future governments detailed control of institutions of higher education, are the most direct threat so far to the integrity of British scholarship and research. Under the terms of the bill, the government would in principle be able to tell individual universities what to teach, and where to concentrate their energies in research. There will no doubt be a great clamour on this score, but in the end the government can use its majority in the House of Commons (where higher education has few friends) to have its way. So the question for 1988 is how to persuade the government itself to change its mind. Will that be possible? We should know by the end of the year ahead. □



Supercollider site selection moves into final round

- Eight states in, but California out
- Final choice up to Energy Department

Washington

A REPORT released last week by the National Academy of Sciences (NAS) has turned eight states into fervent supporters of high-energy physics. On 30 December, NAS issued its shortlist of the eight best-qualified sites for the Superconducting Super Collider (SSC), the \$4,400-million particle accelerator that high-energy physicists and the Reagan administration would like to build.

The sites chosen, one each in Arizona, Colorado, Illinois, Michigan, New York, North Carolina, Tennessee and Texas, include no surprises, but the omission of California from the list was unexpected and has provoked some tentative dissent.

The Department of Energy (DoE) passed on to NAS 36 site applications satisfying the basic needs for a permanent home for SSC, a proton-proton collider capable of collisions at 40 TeV.

To guide NAS in evaluating the proposals, DoE drew up a list of 19 technical criteria which emphasized geological stability and ease of tunnelling, but included such considerations as regional manpower and housing resources and environmental impact. The academy panel was carefully protected from the inevitable political pressure that will be exerted by competing states. Apart from the international prestige, SSC construction will create thousands of jobs, and be a huge boost for any state's economy.

NAS sent its recommendations to the Department of Energy (DoE) on Christmas Eve, expecting that DoE would take time to verify that its selection criteria had been properly considered before releasing the list to the public in early January. But Energy Secretary John Herrington began calling senior political figures from interested states to tell them the news and one, Senator Phil Gramm (Republican, Texas), called a press conference on 29 December to announce the outcome.

Although the NAS report describes strengths and weaknesses of the eight finalists, it is careful not to choose favourites, and gives no specific reasons why other proposals fell short; failure to get on the best-qualified list is attributed by the panel to a combination of doubts rather than particular faults. This has mystified the Californians, whose site near Stockton was thought by many to be strong.

The criteria used by NAS included a provision that time, money and possible litigation involved in meeting local

environmental statutes were to be considered, but although the California proposal had been marred by such worries, the report says that environmental or site-acquisition problems were not insurmountable for any of the bids. Indeed the panel chose the Rochester (New York) site despite receiving many letters of opposition from local residents.

The eight chosen sites all have good geology, and the final selection will probably be influenced by more qualitative factors, especially those in the vaguely worded category of 'regional resources'.

The Illinois bid proposes to incorporate the existing facility at Fermilab into the SSC. This could lower costs substantially, but how important cost will be in DoE's final decision is far from clear.

DoE is sympathetic to the desire of California SSC supporters, and others, to know why their bids failed, and is looking for ways to reveal more without compromising the NAS panel's private deliberations. And although the California SSC committee has publicly stated its disappointment, it is unlikely to protest too vehemently for fear of upsetting the SSC project.

David Lindley

New CERN head

London

THE next director-general of CERN (the European Organisation for Nuclear Research) is to be Carlo Rubbia, joint winner of the 1984 Nobel prize for physics. Rubbia, 53, starts his five-year term on 1 January 1989. A senior physicist at CERN



since 1961, Rubbia at present spends one semester a year in Cambridge, Massachusetts, where he is Higgins professor of physics at Harvard University. The rest of his time is spent mostly at CERN, where he supervises experiments on the proton-antiproton collider. S.L.H.

Shuttle suffers further setback

Washington

THE latest timetable for putting the space shuttle back on active duty has been scrapped after the discovery of serious damage to a solid-fuel booster that was tested on 23 December. The test was initially judged a success, but closer inspection a few days later revealed that a large section of the flexible seal between the rocket casing and the swivelling nozzle had broken away.

The National Aeronautics and Space Administration (NASA) announced on 29 December an indefinite delay to the proposed June 1988 shuttle launch, and until the new problem has been analysed no one is guessing when the shuttle will leave the ground.

Although the joint between the nozzle and the casing was not that blamed for the Challenger accident, engineers at NASA and Morton-Thiokol, where the booster is manufactured, had recognized it as a potential weak spot and had redesigned it. In the latest test, the first full test of the new joint, the nozzle was swivelled to simulate extreme flight conditions.

According to some preliminary reports, the failure of the seal occurred in the last few seconds of the test firing, when the nozzle was moved to the maximum allowed extent, but the design should nevertheless have accommodated the movement. In the best case, the failure will turn out to be the result of a manufacturing flaw, when there might have to be more tests than planned, but the shuttle could still be flown later this summer. In the worst case, if fundamental design problems turn up, the postponement could stretch into next year.

Engineers from both NASA and Morton-Thiokol will study the shell of the failed joint in a matter of days. But until their analysis yields specific pointers to the cause of the failure, NASA will make no comment on the future schedule of testing and eventual flight of the shuttle.

David Lindley & Joseph Palca

The education bill: reform or otherwise?

A SEMINAR on the implications of the Education Reform Bill organized by *Nature* will be held in London on the afternoon of 19 January. The principal speaker will be Sir Mark Richmond, Vice-Chancellor of the University of Manchester and Chairman of the Committee of Vice-Chancellors and Principals. Those wishing to attend should apply for tickets and further information to Mary Sheehan (01-836 6633 extension 2223). □

Soviet academy's restructuring policy causes voting problems

London

THE Soviet Academy of Sciences has been voting in the first of three batches of elections that will bring in some 250 new members and corresponding members. This massive influx is a result of the new "cadre restructuring" policy. From now on, when an academicien reaches 75 years of age, his position will be deemed vacant and a replacement elected. Academicians and corresponding members will have to retire from administrative posts in the academy research network at the age of 70 and other administrators at 65. But although these changes were approved by a general meeting of the academy early last year, the run-up to the elections has evoked discussion on the purpose and role of the academy.

According to *Pravda*, the leadership stated in September that nominees for the vacant places must be under 60 for election as full academicians and under 55 for corresponding members. But when the lists were drawn up in October, many scientists over 60 were included.

A number of academy members think the age limit contravenes the academy's statutes. Others feel that the limit would be detrimental to science. In particular, many stress that inherent conflict in the double role of the academy — as an association of eminent scientists, with membership bestowed as the crowning honour of a brilliant career, and as the principal coordinator and organizer of research in the Soviet Union.

There has been less opposition to the compulsory retirement of superannuated administrators (12 senior academicians have already resigned under this rule) than on the upper age limit on election. Almost a third of the existing members of the academy, it is argued, were elected only when past 60, and some deserving scientists over 60 will now be excluded for life from an honour they would otherwise almost certainly have received in the next few years.

Moreover, it would not be simple to devise an interim solution that would allow, for the next decade or so, the existence of supernumerary members drawn from deserving but now over-age scientists. For academy membership involves not merely a title of honour but also a salary and fringe benefits. Some scientists now question this system, which although not unique to the Soviet Union is out of step with general world practice. If scientists are paid by their work, it is argued, rather than for their title, they will be encouraged to produce tangible results.

Abolishing academic salaries and benefits, moreover, would help to put

an end to the current 'back-door' admissions. Scientific administrators and directors of institutes are frequently elected to the academy as a matter of routine. And such members not only have little to offer in terms of scientific achievement, but have also, it is argued, introduced an "alien bureaucratic style".

The new rules on compulsory retirement will inevitably entail a major influx of new administrative personnel into the academy institutes, making it all the more important to end the practice of routinely electing such officials to academy membership. Suggested measures include stricter assessment of the scientific work of those proposed for election, greater 'democratization' of the elections (with corresponding as well as full members of the academy being allowed to vote) and a

special arrangement for the Siberian, Far Eastern and Ural branches, allowing those who work there to be made 'candidate members' of the branch.

In spite of numerous calls to "restructure" personnel policy, the old scheme of routine promotions still to some extent persists. One proposal voted on last month was the nomination of Aleksander Zhuchenko to be director of the academy's Institute of General Genetics. Zhuchenko, originally an agronomist, rose to be president of the Moldavian Academy of Science, largely due to the patronage of the now-disgraced first secretary of Moldavia, Ivan Bodul. Since Bodul's fall, Zhuchenko has come under severe attack for the agricultural policy he urged on the Moldavian authorities, which has led to the erosion and salination of considerable tracts of land. The fact that he could nevertheless be proposed to be director of one of the academy's most important institutes indicates, perhaps, the scale of the problem.

Vera Rich

West Germany looks for the neural-network connection

Munich

Anxious about losing the race to build neural-network computers, West Germany is beginning a three-year project in January 1988. 'Information processing in neural architectures' will be funded with DM3 million in 1988 and possibly much more in the future.

The project represents an "extraordinary effort" by the West German Research and Technology Ministry (BMFT), says Rolf Eckmiller of the University of Düsseldorf, one of the project's beneficiaries. Nine groups will be funded, with three or four new positions at the universities of Düsseldorf and Bochum. This stands to make the *Land* of North Rhine-Westphalia, where these universities are located, "not only the German centre but the European centre as well" for research on neural networks, says Eckmiller.

The goals of the project are twofold: to transfer knowledge about how the brain functions to the design and programming of new parallel computers; and to develop intelligent robots that will be able, using neural processing, to "see, hear, write, speak" and otherwise respond to their environment.

Despite the excitement in US, Japanese and European laboratories about the commercial potential of neural-network computers, BMFT is not pushing West German researchers exclusively towards applications. "The BMFT's goal is to support basic research", according to one project leader, Klaus Schulten of the Technical University of Munich. "They're not pressing us to come up with something

that can be sold within five years." International competition will be stiff. According to one estimate, the United States is far in advance, with roughly 2,500–3,000 researchers working in the field. Japan follows with 100–500 people. In the European Economic Community (EEC), there are 500–1,000 researchers, 100–300 of them in West Germany.

The EEC has its own programme, entitled BRAIN (Basic Research in Adaptive Intelligence), to support research in neural networks. But its funds are not awarded on scientific merit alone, priority being given to groups that collaborate internationally. Funds are also available through another European programme, ESPRIT II.

The interdisciplinary nature of neural-network research makes it unpopular in the more rigid of the West German universities. In fact, Schulten moves to a US university in 1988 where he will not be "isolated and looked down upon". He will take his entire research group with him when he goes. The group is trying to programme a robot to learn from its own mistakes, for example in distinguishing and manipulating blocks of different shapes and colours.

For Research and Technology Minister Heinz Riesenhuber, the fact that good people are leaving provides a compliment to the quality of West German research — but a compliment that is not entirely welcome. In a recent television interview he said that one goal of the new project is to keep good people in West Germany.

Steven Dickman

US Congress acts on fears of nuclear plane plutonium disaster

Washington & Tokyo

FEAR of a major nuclear disaster has driven a US Senate panel to take the unusual step of sending back to President Ronald Reagan for renegotiation the US-Japan nuclear agreement signed in November. The pact calls for plutonium to be flown from Europe to Japan by the polar route, raising fears that an accident could contaminate large areas of Alaska with deadly plutonium. Renegotiation will not be easy: a separate amendment to the budget bill requires the safety of plutonium storage casks to be tested in a plane crash before any flights are authorized.

Japan wants the plutonium, produced as a by-product in its light-water reactors, to use as a nuclear fuel. But to do so requires considerable international cooperation. First, Japan must send its spent fuel to France or the United Kingdom for the plutonium to be extracted and purified. Japan has only a tiny reprocessing plant of its own. And, second, Japan must obtain US consent because the fuel was sold to Japan under stringent US nuclear non-proliferation regulations.

The biggest problem is how to return the plutonium from Europe. Only 15 lb is needed for a nuclear bomb so security has to be strict. An earlier attempt to send plutonium by sea from France turned out to be enormously expensive when, after two years of delay, an escort of two warships and continuous satellite surveillance had to be provided to satisfy the United States. The new US-Japan agreement would provide a cheaper route. Plutonium would be shipped by air, probably by the polar route via Anchorage (see *Nature* 330, 102; 1987). About one flight every two weeks, each carrying 500 lb of plutonium, would be required.

To avoid risk, the agreement specifies that the plutonium storage casks must be crash-proof. But, as Senator Alan Cranston, a member of the Senate Foreign Relations Committee which voted down the agreement put it, "the crash-proof cask is not yet invented".

According to reports gathered by the Nuclear Control Institute, a Washington-based non-partisan policy institute, casks developed in the United States in co-operation with Japan, and in France, have failed miserably when slammed into a hard target at 250 knots. And as it is far from clear that the effects of a plane crash can easily be simulated, Congress has now specified that casks must be subjected to a real crash test unless an outside panel of experts agrees that a laboratory test of a worst-case accident is feasible.

At \$25 million, the crash of an old

Boeing 747, full of test equipment, will not come cheaply. But supporters of the amendment say that it is a small price to pay for the security of knowing that a crash would not scatter plutonium over Alaska. If things really went wrong, cleaning up after a crash would be prohibitively expensive. When a military plane carrying nuclear warheads went down at Palomares, Spain, in 1966, more than \$500 million was spent on removing plutonium-contaminated soil. A second accident at Thule, Greenland, in 1968 cost \$300 million and required the removal of 150 mil-

lion gallons of snow, ice and water.

Rejection of the agreement by the Senate committee has already received backing in a similar move by the House of Representatives Foreign Affairs Committee. Both argue that the agreement does not meet the terms of the Nuclear Non-Proliferation Act. President Reagan is now faced with some difficult choices. He has to decide whether to renegotiate the agreement, exempt it from the non-proliferation act and send it back to Congress to see if he can win majority support in both Houses, or insist on it becoming law and risk a fight with Congress. The level of opposition now apparent in Congress suggests that he will take either of the latter two steps at his peril. **Alun Anderson & David Swinbanks**

Enrolment mix raises ethnic issue at Berkeley, California

Berkeley

ASIAN-AMERICAN groups are in dispute with the University of California (UC) at Berkeley over admission policies which they claim discriminate against Asian-American applicants. That charge may seem implausible — Asian-Americans already take 40 per cent of undergraduate places in engineering and chemistry, and 28 per cent of total freshman class places. But Asian-Americans argue that if the admission process were unbiased, their admission rates would be higher still.

The remarkable academic performance of Asian-Americans is not restricted to Berkeley. Overall, Asian-Americans make up just 2 per cent of the US population. But at top universities they are consistently 'over-represented', taking 16 per cent of freshman places at Stanford, 14 per cent at Harvard and 20 per cent at the Massachusetts Institute of Technology.

The question of discrimination has also been raised elsewhere in California, but nowhere has the problem attracted so much attention as at Berkeley. Figures show that the overall admission rates (percentage of applicants admitted) are slightly lower for Asian-Americans than for comparable white applicants.

At Berkeley last year, nearly 2,000 undergraduate applicants with top examination scores had to be turned away. Half of Berkeley's entrants are chosen on test scores alone and most of the remainder on evaluation of mathematics and foreign language background, English proficiency and a written essay. Under-represented minorities (blacks, Hispanics and American-Indians), recruited athletes, handicapped and rural students, are guaranteed admission if they meet basic UC eligibility requirements.

Asian-Americans claim that the way credit is given for foreign languages lowers

their chances of admission. Students who have grown up speaking an Asian language cannot use that skill to their advantage because the US standardized achievement tests do not include tests in Asian languages. The university counters by arguing that the difference in admission rates can be accounted for by the under-representation of Asians among the recruited athletes and rural students guaranteed admission. But the university now intends to give full language credit to students who have immigrated from non-English speaking countries.

The graduate admissions process may be more difficult to evaluate, says Eugene Cota-Robles, assistant UC vice president for academic planning. Asians comprise 8 per cent of UC graduate students, and are clustered in the professional schools and science and engineering departments. Cota-Robles says that some apparent discrepancies between admission rates for Asian and white applicants to UC graduate programmes have actually resulted from a skewing of Asian applicants toward highly competitive programmes such as medicine, law, business and engineering.

Nevertheless, further examination of graduate admissions on the Berkeley campus has shown that Asian admission rates lag behind white rates by 16 percentage points in some graduate departments. The reasons for the discrepancies are as yet unclear, but the dean of the graduate division has launched an inquiry into admissions policy in those departments.

Berkeley chancellor Ira Michael Heyman has also set up a campus task force, made up of students, staff, faculty and members of the Asian community, to examine all of the opportunities and disadvantages faced by Asian students at Berkeley. **Marcia Barinaga**

Closer ties in view for Soviet Union and India

Bangalore

CLOSER scientific ties between India and the Soviet Union are being created in the follow-up to the long-term programme of cooperation signed by the Soviet General Secretary, Mikhail Gorbachev, and the Indian Prime Minister, Rajiv Gandhi, in July 1987. Eight areas have been named for potential collaboration, including space research and nuclear technology.

As a part of the agreement, a Soviet Proton booster will orbit an Indian-built Earth observation spacecraft, IRS-1A, early next year, and its successor, the IRS-IB, in a few years time.

The Soviet Union has also agreed to loft an Indian-built microwave sensor payload on board a Soviet spacecraft sometime early next decade. More important, the Soviet leader has offered to set up a space research centre in India to help developing countries to obtain expertise.

Nuclear collaboration is proving more difficult. Controversy has arisen over the Soviet offer of a nuclear plant to India. The Soviet Union has been pressing India to throw open its nuclear facilities for inspection by the International Atomic Energy Authority. And although India's minister of state for science and technology has gone on record saying that India is actively considering the Soviet nuclear offer, there are others in India, including scientists and several engineering companies, who are suspicious that the offer may be an attempt to bridge the Soviet trade gap with India.

Many scientists are reluctant to proceed with nuclear power development that relies on foreign technology. Although eager to learn about uranium-enrichment technology, they feel that the best course for India would be to start off by building a low-level enrichment facility, relying on its own technology and facilities, rather than depending on an alien supplier, such as the Soviet Union, that would probably impose stringent conditions on the operation of any facility that it helped build.

Several Indian engineering companies, including Tata, L&T and Walchandnager, have recently expanded their operations to meet the demand for India's nuclear power programme and are now worried that the Soviet entry into the Indian nuclear power programme would harm them.

Radhakrishna Rao

Vienna Biozentrum: correction

Rudolf Schweyen, head of the Institute of Microbiology and Genetics at the University of Vienna, is to be a senior faculty member of the Vienna Biozentrum (see *Nature* 330, 686; 1987). Erwin Wagner will join the Institute for Molecular Pathology. □

Fears of recession complicate outlook in United States

Washington

Although the general health of the US scientific enterprise remains excellent, the events of 1987 made it clear that US scientists — and indeed all US society — can no longer afford to ignore the outside world. The stock market's tumble underscored the global interdependence of the US and other economies.

Congress now aims to support research that can help "economic competitiveness", a phrase that became fashionable in 1987's grant applications and is likely to be worn by everyone in 1988. Justifying spending on pure science, especially on expensive items such as the Superconducting Super Collider (SSC), will be increasingly difficult.

The urgent need to trim the budget deficit may not require immediate sacrifices. So far, the Reagan administration and Congress have chosen to delay or slow down major new scientific projects, rather than eliminate them. Final decisions may well be postponed until the new administration takes over in 1989. But if the economy recovers, lost time can be made up.

Private universities weathered the crash of 1987 with the comfortable feeling that their losses were much smaller than the gains that they had chalked up earlier. In 1988 they may not be so fortunate. They, and others lacking massive federal support, will be most at risk should there be an economic downturn.

As economic forces will alter how research is conducted, so too the recently signed treaty eliminating intermediate and shorter-range nuclear missiles could signify a change for military research. Budgetary and political forces seem already to have stopped the Air Force's expensive and potentially destabilizing anti-satellite programme. But another potential victim of arms control, the Strategic Defense Initiative, seems destined to survive as President Reagan's pet project, even if Congress will not provide all the funds he would like.

The civilian space programme remains grounded as NASA (National Aeronautics and Space Administration) works towards returning the shuttle to operational status. Unhappily, problems with the latest shuttle motor test have caused a further postponement of flight resumption (see page 2). NASA is also pushing hard for its \$14,000 million manned space station, and the first round of contracts to build it have been awarded. But its high cost remains a liability.

The global effects of atmospheric pollution served as another example for the United States that unilateral action is not

always the answer. International talks led to success last year in the signing of the Montreal Protocol to reduce ozone-destroying chemicals. By contrast, efforts to reduce pollutants causing acid rain are mired in nationalistic squabbling between the United States and Canada.

The debate over whether to build SSC, a \$4,400 million proton-proton collider, has become tangled in claims that it will produce innumerable technological spin-offs making it an investment in the country's future. But much of the clamour comes from states fighting to provide the site (see p.2). So far, the Department of Energy has moved forward on schedule: a shortlist of best-qualified sites has been released, with a selection in June.

The unexpected discovery of high-temperature superconductivity put some glamour back into science in 1987, but commercial applications are a long way off. Those politicians who advocated the immediate and generous spending to bring the new superconductors to a state of industrial use are beginning to see for themselves that basic and applied research are not neatly divisible.

AIDS will continue to dominate biomedical research in 1988. With an effective vaccine still far off, the focus will be on treatment and education. As the disease spreads, the strain on the health-care and life-insurance systems will begin to show. Federal coordination of efforts to combat AIDS is likely to remain disjointed and ideologically driven, but federal money should continue to flow.

Biotechnology companies weathered the market buffeting well, although their high cash reserves and rock-bottom prices briefly made them attractive take-over targets. Despite the crash, biotechnology became profitable in 1987 with the approval of tissue plasminogen activator (TPA), the first biotechnology drug with a substantial market. But patent disputes continue to be a drain on resources. Negotiating the maze of regulations governing biotechnology will continue to put a strain on the industry, but environmental release experiments went forward last year despite critics, and gene therapy experiments may take place this year.

A project to map and sequence the human genome has started, although whether it will be directed by a central authority remains uncertain. The scientific value of a complete sequence is unquestioned, but political support for the project may derive from fear that if the United States does not complete the sequence, the Japanese will.

Alun Anderson, Carol Ezzell,
David Lindley & Joseph Palca

West Germany's 'Greens' fade as opponents go environmentalist

Munich

THE latest scandal involving the West German nuclear power industry brought Environment Minister Klaus Töpfer's 1987 Christmas vacation to a premature end. The uproar may set the tone for the year to come.

Töpfer took office in April following the election of his predecessor and fellow Christian Democrat (CDU) Walter Wallmann as Minister-President of Hesse. Wallmann's victory in the former Social Democratic (SPD) bastion of Hesse seemed to solidify the pro-nuclear conservative government of Helmut Kohl.

But Töpfer found himself in a hot seat from the start. The current scandal, front-page news in West Germany since just before Christmas, involves the nuclear waste transport company Transnuklear, based in the Hessian city of Hanau. Transnuklear has been charged with illegally shipping more than 1,700 barrels of radioactive waste, some 321 of them known to contain traces of plutonium, under false labels to temporary storage sites in West Germany. The barrels had first been shipped to a Belgian nuclear energy centre in the city of Mol, where the waste was meant to be processed.

A bribery scandal involving Transnuklear and two companies in Mol has added fuel to the fire. The head of Transnuklear's 'radioactive substances' division admitted his guilt in the bribery case in early December and committed suicide while in police custody several days later. Töpfer will return from vacation before the new year to determine if the bribery is connected to flouting safety regulations.

The issue of regulation of the Hanau nuclear companies (Alkem, Nukem and Transnuklear) was an Achilles heel for the Greens in 1987. Their coalition with the SPD in Hesse ended in February because of a dispute over the processing of plutonium at Alkem. Despite a strong showing at the polls, the Greens were shut out of the government by Wallmann's CDU in the 5 April election.

Since then, the Greens have almost disappeared from the national political stage. The party has splintered into at least three factions: fundamentalists, 'realists' and those somewhere in between. Their public sniping has grown so intense that local Green Party members have been travelling to Bonn and begging their leaders to get back to business. At the same time, the other parties have begun to take more 'environment-friendly' positions, which has weakened the Greens yet further. Unless the Greens can win over the SPD, which views them with barely concealed repugnance, they seem destined to remain

in the shadows this year.

German environmentalists, the Greens among them, watched with dismay in 1987 as important precedents were set in biotechnology. First, biologists released into the environment nitrogen-fixing bacteria in which an antibiotic resistance gene had been inserted. Then, the *Land of Hesse* gave pharmaceutical giant Hoechst a licence to produce synthetic human insulin from genetically manipulated bacteria. The Greens had held up the licensing procedure for two years. Boehringer Ingelheim and others await the chance to use genetic engineering in production, and there is apparently nobody strong enough to stop them. The *Land of Bavaria* should continue to make headlines in 1988 with its controversial AIDS (acquired immune deficiency syndrome) policy. In Bavaria, all non-European Community residents must be tested for the human immunodeficiency virus (HIV). Carriers have to leave.

Bavaria's conservative government has made enemies among coalition partners and opponents alike with its hard-line policy. It has also run into some noteworthy difficulties: the Bavarian Interior Ministry ruled in August that immigrant monks must be tested despite their vow of celibacy. But three Soviet technicians at an Augsburg machinery company refused to be tested and were eventually excused from the test. Foreign residents of Bavaria who arrived before 1 June 1987 will be tested in 1988.

In 1987, the West German government reached important agreements with several Eastern and Western European neighbours. Perhaps the most significant among these were those on environmental and scientific exchange with East Germany and the decision to go ahead with the Ariane 5 rocket, the Columbus space station and the Hermes space shuttle projects of the European Space Agency (ESA). The former reflects a historic relaxation of tensions between the two German states; the latter illustrates the health of the relationship between West Germany and France.

But the price will be high. Virtually any exchange with East Germany winds up supported by West German currency. The space programme threatens to drain thousands of millions of deutschmarks from the Research and Technology Ministry.

Barring surprises, there will be no Bundestag (federal parliament) and only two *Landtag* (state parliament) elections in 1988. But given the environmental and international pressures on the government, it promises to be quite a lively year.

Steven Dickman

Prospects facing as US pressure

Tokyo

WILL 1988 be the year in which the Japanese government finally commits substantial new funds to basic scientific research? The conditions are ripe. Events in the past year have awakened the Japanese public to the importance of pure science: high-temperature superconductors, AIDS (acquired immune deficiency syndrome), a supernova, Susumu Tonegawa's Nobel prize, and the Human Frontiers Science Programme have made basic science a hot public issue. Added to this, the United States is putting political pressure on Japan to open up its research system to the West and the issue of basic scientific research is threatening to run into yet another 'trade dispute' between the two economic giants.

Overall, 1987 was a memorable year for Japanese science. The year began with Tokyo University's confirmation of high-temperature superconductivity in a copper-oxide ceramic which sent Japanese physicists scurrying to their laboratories while superconductor-related shares soared on the Tokyo Stock Exchange. And during the first half of the year, hardly a week went by without the announcement of another 'breakthrough' in the development of the new superconductors.

Physics again hit the headlines in March when a group at Tokyo University detected neutrinos from the supernova in the Large Magellanic Cloud and in November when the Institute of Space and Astronautical Sciences (ISAS) announced detection of X-rays from the same supernova (see *Nature* 330, 230-285; 1987).

But physics was not the only scientific field to grab the public's attention. The death of a young woman from AIDS in January shocked the nation. The Ministry of Health and Welfare quickly drafted legislation to force doctors to disclose cases of the disease and the official count of patients and carriers, while still low compared to the United States, Europe and Africa, has risen dramatically.

Susumu Tonegawa's recent award of the Nobel prize for medicine and his bitter criticism of Japan's research system has also brought basic science into the limelight. And even the Human Frontiers Science Programme, which has been bogged down in a seemingly endless feasibility study, has played a role in keeping basic science in the public eye.

The United States is demanding reciprocal access to Japan's basic science and technology. And, in response, the Science and Technology Agency's latest white paper calls for more support for basic research and increased "internationalization" of Japan's science.

All the science-related ministries have

Japan in 1988 increases

leapt on the basic science band wagon with requests for funds for basic research in fiscal year 1988. The Health and Welfare Ministry has requested more than a thousand million yen (about \$10 million) for AIDS research, an order of magnitude up on last year's request. Both the Ministry of International Trade and Industry and the Science and Technology Agency have asked for several thousand million yen for superconductor research. The agency and the Ministry of Education, Science and Culture have also requested almost a thousand million yen to invite several hundred foreign postdoctoral fellows to Japan.

But although these budget requests for the fiscal year 1988 were in large part accepted by the Ministry of Finance during the final round of negotiations at the close of last year, they barely touch the surface of the problem. Japan now spends close to ¥ 10 million million yen (\$80,000 million) a year on research and development, but 80 per cent of the research is funded by the private sector, most of it is applied, and each year the gap widens between government and private sector funding for research.

But political pressure is mounting both within and outside the country for a substantial improvement in Japan's basic research system. The issue will be central to renegotiations this week in Tokyo of the US-Japan Science and Technology Agreement and will probably come up for discussion when newly elected Prime Minister Noboru Takeshita visits the United States next week. And basic science will again be on the agenda in March at the Organisation for Economic Cooperation and Development (OECD) at a meeting of which the United States intends to propose a 'rule' that OECD member nations, including Japan, should contribute to basic research in proportion to their economic strength.

The United States is also demanding access to Japan's industrial research system and is pressing Japan to institute legal controls to protect technological secrets that could have military applications. But these demands will be hard to meet, further increasing the pressure for Japan to improve and 'internationalize' the basic research system in universities and government research institutes throughout the country.

The Ministry of Finance will no doubt fiercely oppose any requests that require a substantial outlay of funds of the order of thousands of millions of dollars. But the ministry may find that 1988 is the year in which it has to heed these calls for change.

David Swinbanks

Indian gloom over drought lifted by 'big science' projects

New Delhi

THE award of US\$290 million as interim relief to Bhopal gas victims three years after the tragedy is the biggest end-of-year news for Indians, but celebrations must wait until Union Carbide actually pays.

The 17 December ruling by the Bhopal district judge, Mr M. W. Deo, is the first glimmer of hope for the relatives of the 2,000 dead and for the 200,000 surviving victims of the poisonous methylisocyanate gas leak from the Carbide pesticide plant in December 1984. Union Carbide has described the court order as "amounting to award of damages without trial" and is threatening to appeal on the grounds that interim relief cannot be paid while liability is in dispute.

The Indian government is suing Union Carbide for \$3,000 million in damages, but the US company maintains that it is not responsible for the tragedy which, it says, was caused by employee sabotage. India has rejected an offer of an out-of-court settlement of \$600 million.

The good news, however, has been buried in problems caused by mounting inflation and the current drought. The outlook for science in 1988 depends on how Prime Minister Rajiv Gandhi copes with economic problems caused by the failure of the monsoon. After priding itself on its self-sufficiency in food, India will have to import rice, edible oil and probably wheat. Gandhi's priority is to use available knowledge to fight the drought and scientific agencies have been asked to pool their knowledge and produce a blueprint early next year.

The gloom cast by the drought will not, however, hamper projects on 'big science': superconductivity, metallic glass and the world's biggest metre-wavelength radio telescope. Massive computerization is on the cards, starting with railways and banks. A satellite-based computerized network called NICNET will next year, link every administrative unit at the district level with state and central ministries.

Gandhi has said that socialism does not work in India and has accused public-sector enterprises of draining the nation's wealth. The private sector will come to the fore, and more projects will be given to industries. Major government spending will be defence-related: missiles, the main battle tank (MBT) and light combat aircraft.

When the prototype of MBT is ready, a production version will enter user trials. Work on a missile test range begins next year where four different guided missiles already developed will be tested. The new year will also see the test flight of India's first re-entry vehicle.

For the Indian Space Research Organi-

zation (ISRO) the new year will be hectic. The launch of the augmented satellite launch vehicle is scheduled for February. It fell into the sea on its maiden flight last year and the fault was never identified, but ISRO is hoping it will fly. In June, the Soviet Union will orbit an ISRO-built remote-sensing satellite for a fee of \$6 million and around that time Ariane will put into orbit INSAT-1C, a communications satellite built by Ford Aerospace to replace INSAT-1B.

The year ahead will see greater application of relevant science for improving the basic quality of life of Indians. Gandhi has replaced Professor M. G. K. Menon with Sam Pitroda, a telecommunications wizard, as chief of technology missions on drinking water, immunization, communications, oil seed and literacy.

The government took a gamble last year in bringing nuclear power generation under public control in the hope that banks and private individuals will invest in the company. Indians' attitude to nuclear energy will become clear when the Nuclear Power Corporation starts selling its bonds and shares next year.

The government is pleased with the results of the phase one trial of an Indian-made birth control vaccine and large-scale trials on women will begin next year. Having introduced into the market a plant-based drug (for hypertension), the government is reviving research on remedies based on ancient Indian systems. Two herbal drugs for jaundice and filariasis will enter trial.

Because of extensive deforestation, a total ban on tree felling will be enforced, though Gandhi is unable to stop clearing of forests for hydroelectric projects, some of them with World Bank aid. The Soviet Union will begin construction of the massive Tehri Dam in the Himalayas against local opposition.

The year ahead will see launching of new collaborative ventures with the Soviet Union and the United States. The Soviet Union, which has leased two nuclear submarines to India, will sell two of its nuclear power plants. India has agreed in principle to a Soviet proposal for an international space centre in India for astronaut training and joint space launches.

At long last the US supercomputer will arrive next year under conditions that some say will impinge on national sovereignty. The Indo-US agreement on testing of genetically engineered vaccines remains in limbo. The agreement has not been scrapped, but it is doubtful if any vaccine will be tested in India unless already approved for use in the United States.

K.S.Jayaraman

France puts trust in prestige projects

Paris

IF a single phrase had to be found to summarize the mood of French science in 1987, it could well be 'keeping up appearances'. France has said 'oui' to many of the expensive, European big-science projects, to which Britain has said 'no', yet France is no better prepared, financially, to take scientific leadership in Europe than is Britain. Meanwhile, behind the scenes, it has been a year of political turmoil, cuts in basic science and budgetary sleight of hand.

The year began without a science minister in the still-new conservative government. Alain Devaquet was forced to resign just before Christmas 1986 after his disastrous proposals for university reform brought protesters on to the streets, leaving one student dead. His successor, Jacques Valade, was saddled with the task of sorting out the imbroglio in France's major public-sector research body, the Centre National de la Recherche Scientifique, which has a budget of £850 million. Its scientific appointments committee had been illegally dismissed by Devaquet, freezing recruitments and promotions until a high court ruling reversed the ex-minister's decision in the middle of the year.

France, like other European countries, is feeling the technological dominance of the United States and Japan and knows that it has to invest more in research and development to keep up. But the government still lacks a coherent science policy. A large part of the science budget is now allocated by the Industry Minister and Minister of Finance, who, for 1988, gave an invisible, but on paper, large (£50 million) boost to research and development in the form of tax incentives, despite a lack of evidence that it has worked elsewhere.

Even if all is not well at home, France has tried to project a decisive and forward-looking attitude by backing all the major European collaborative research facilities, including the synchrotron and the CERN particle accelerator, as well as the European Community's science and technology programmes.

France's chief success, though, has been the European Space Agency's long-term programme, which was approved by all member states except Britain. France will be responsible for 45 per cent of the cost of the Ariane 5 heavy-lift rocket launcher and managed to push through its pet, the Hermes shuttle plane, largely because Britain's revolutionary horizontal take-off and landing craft (HOTOL) could not be demonstrated to work.

French basic science achievements were less evident, but Jean-Marie Lehn's share of the 1987 Nobel prize for chemistry came as a welcome tonic.

Peter Coles

Soviet *glasnost*, *perestroika* and democratization

London

DURING 1987, *glasnost* focused in particular on environment and health. Statistics were published on formerly sensitive issues such as occupational accidents and diseases, hepatitis and infant mortality (which for some regions is admitted to be comparable with developing countries).

The threat of AIDS has been recognized, and a doctor claiming to have an anti-AIDS vaccine received media coverage, although not the support of the Academy of Medical Sciences. Environmental campaigns have flourished and many 'informal' groups of young environmentalists have been formed. Environment-related demonstrations in Erevan and Minsk have, for the first time, been reported in the official media, and some concrete decisions were made in 1987 — the proposed Daugavpils hydroelectric plant and plans for two of a projected twelve nuclear power/heating stations for Soviet cities were abandoned.

But discussion (and, political climate permitting, demonstrations) will continue in 1988, particularly where offending enterprises ignore official directives and enforcement orders, as in Armenia and Kirgizia. Moreover, Soviet planners will have to reconcile the conflicting demands of different groups. Thus the decision, in 1986, to halt the scheme to divert the Siberian rivers southwards has exacerbated the water problems of the Central Asian republics, and generated new campaigns.

Glasnost, however, does not operate uniformly. Many ministries and government departments still pay it only lip-service, and a new word, *nedoglasnost* (unachieved openness) has been coined by disgruntled journalists. Some unclassified documents (such as plans for an anti-narcotics campaign) have been improperly withheld from the media; other information (such as the three fatal accidents at the Chernobyl site in 1987) is mentioned only in the specialist press. During the past year, hints of a 'Challenger-scale' near-disaster (apparently a launch-pad abort in 1983) have appeared in the Soviet media, but according to Vitaly Korotych, editor of *Cgonek*, there has been considerable pressure from the *Glavkosmos* space administration to suppress the story.

On formally secret issues, there has been some progress, both in the military field and on the secrecy restrictions on emigration. Some prominent *refusniks* have been granted permission to leave the country, including Aleksandr Lerner, Aleksandr Ioffe and Irina and Viktor Brailovskii. But these have been *ad hoc* decisions, and Mr Gorbachev has made it clear that an across-the-board change of

policy is not contemplated. The International Federation of Scientists for Soviet Refusniks is now preparing a new campaign calling on the Soviet authorities to draw up, implement and publicize a clear policy on the secrecy issue, with a maximum cooling-off period that corresponds to the current rapid rate of scientific progress.

Perestroika, now that the most notoriously slack institutes have been closed, focuses on shedding excess personnel and restructuring research to make it more 'relevant' to industry and agriculture. Neither is easy. There is a strong lobby within the Academy of Sciences against compulsory retirement and the new upper age-limit on appointments, which may well have a knock-on effect. 'Relevance' is hard to ensure; the planned 'research-production associations' intended to break bureaucratic deadlocks are slow to get started, and even show-pieces of relevance such as the Paton Welding Institute of the Ukrainian Academy of Sciences seem to have lost some of their media appeal. And the supreme example of 'relevance' in the space programme — the promise of zero-gravity technology aboard a permanently manned space-station — which for three decades has been used to still criticism of the cost of space, is itself in doubt. According to Vladimir Shatalov, head of cosmonaut training, new developments in terrestrial technology since such a station was first planned have made the planned 'weightless' workbench largely unnecessary.

'Democratization' in science means the election of institute and laboratory directors, and is, in fact, lagging behind the country at large, as posts in academy institutes are decided by academy members only, without the participation of institute staff (who are pressing for change). And there are disturbing reports from Novosibirsk of a candidate for an institute directorship falling victim to an unfounded smear campaign by the Russian-nationalist antisemitic *Pamyat* (memory) association.

What will 1988 hold for the Soviet Union? Primarily a continuation of Gorbachev's threefold reforms. It will also, almost certainly, involve a propaganda drive to promote the marxist-leninist *weltanschauung* (world philosophy), to counter the revival of Islam in the southern republics and the forthcoming millennium of the 'Baptism of Rus'. Science will undoubtedly play a major part in this campaign, and full advantage will be taken of the Mars launch 'window' which conveniently coincides with the Christian celebrations.

Vera Rich

Value-for-money yardstick now dominates UK science decisions

London

BRITISH science stumbled on through 1987, despairing at the perennial shortage of funds but alarmed and indignant at suggested solutions. The return to power, in June, of Mrs Margaret Thatcher's Conservative government for a third term signalled little respite from the financial hardships of the past. Instead, 1988 will see the start of the evolution of a new framework for the management of science, with industrialists wielding greater influence in the distribution of the science budget, and a concentration of research in fewer institutions.

For the universities, the coming year will be dominated by the passage through parliament of the Education Reform Bill of Mr Kenneth Baker, Secretary of State for Education and Science. Since its publication last November, initially muted protests at the bill's proposals on university funding and academic tenure have steadily grown into a full-throated clamour of opposition. The next few months will see spirited lobbying in the House of Lords, where the bill is now being debated, to secure provisions for guaranteeing academic freedom in the absence of tenure and loosening the direct hold of education secretaries on the distribution of funds between institutions.

Academic science found an almost unanimous voice in the condemnation of proposals, published in July by the Advisory Board for the Research Councils, to divide higher education institutions into three types, with varying levels of research and teaching. The government is expected to announce its position on the proposals by the early summer. Whether the institutions will be classified as suggested by the report is less certain than the support the government will give the formation of interdisciplinary research centres attached to institutions. The Science and Engineering Research Council has already started this process.

The nature of the work carried out in the university research centres will be determined to a large degree by two new bodies announced in 1987, one, ACOST (Advisory Council on Science and Technology), to advise the government exclusively on areas of scientific priority, and the other, CEST (Centre for Exploitation of Science and Technology), to alert industry and the government to emerging areas of strategically important technology. The voice of the businessman resounds loudly in both.

In January, Sir Frank Layfield recommended that the construction of a pressurized water reactor at Sizewell, in Suffolk, be allowed. Official government approval

soon followed and work started in July, reviving Britain's nuclear power programme, on ice for nearly a decade. Approval for a similar station at Hinkley Point, in Somerset, is expected this year.

The post-election government reshuffle saw the ejection of Mr Geoffrey Pattie from his position as Minister for Information Technology at the Department of Trade and Industry (DTI). Pattie was a keen advocate of an expanded space programme, recommending a doubling of the national space budget to £200 million. The British National Space Centre (BNSC), formed in 1984 to coordinate the nation's space effort, had in September 1986, at the government's request, produced a 15-year plan outlining options for British involvement in space, requiring extra sums ranging from £90 million to £200 million. Last July, Thatcher announced that no more money was available for space. A few weeks later, BNSC's director-general, Mr Roy Gibson, resigned, having been refused an extra £7 million to keep open British options on various European collaborative programmes. Since then, Britain has isolated itself from its fellow members of the European Space Agency (ESA) by refusing to take part in expensive long-term infrastructure projects and vetoing an increase in the mandatory science budget. The government wants to see private industry put more money into space research; 1988 will see BNSC become more orientated to the needs and interests of industry.

As at ESA, British involvement with two other international ventures was plagued by uncertainty, although ultimately with happier outcomes. Continued membership of CERN, the European Organisation for Nuclear Research, was eventually agreed in December, after months of nail-biting in the high-energy physics community. Britain managed to squeeze aboard the European Synchrotron Radiation Facility at a subscription of 10 per cent — less than the going rate.

A follow-on to the Alvey programme of advanced information technology will be announced in 1988. The new programme will be closely tied with the European Community's ESPRIT programme, and the government's contribution will be far less than the £425 million recommended by a committee whose findings have been with the department for more than a year. An internal DTI review may succeed in disentangling from red tape the government's much-vaunted LINK programme for which £210 million is allegedly available to encourage collaboration between industry and universities.

Simon Hadlington

Astronomy looking up in Australia

Sydney

"MAKE it pay" has been the government's message to science in 1987 as Australia reacts to changing patterns of international technological leadership, and sees the wealth that technology can create.

According to Barry Jones, whose own change in title — from head of the Department of Science to Minister for Science and Small Business within the Department of Industry, Trade and Commerce — symbolizes the change in thinking, funding for scientific research will depend increasingly on how much progress scientists can make in finding partners in industry.

Jones believes that one of this year's biggest issues will be the question of concentration of resources. Is it right to spread resources thinly and equitably across all areas of research or is it more sensible to concentrate resources to create a 'critical' mass in a few strategic areas?

The question will be high on the agenda of the Australian Research Council (ARC), a new body which will be responsible for most funding of university research, and answerable to the new Minister for Education, Employment and Training, John Dawkins. ARC's aims will be to direct research resources in the service of national goals rather than purely on the criterion of academic excellence. Not until Dawkins appoints the council's chairman and staff will it be apparent whether short- or long-term goals will dominate policy.

A scientific breakthrough in the handling of coal is one that Barry Jones boasts of as the kind of thing Australia needs. Announced a few weeks ago, it involves liquefying coal into a fine slurry that can be handled as easily as other liquid fuels. Jones hopes it will earn A\$300 million a year in exports while revitalizing the coal industry and cutting pollution.

The coming year is to be one of national celebration: 1988 is the bicentenary of the European settlement of Australia. Amid the festivities will come major boosts for Australia's traditional strength in astronomy: the commissioning of the "Australia Telescope", the largest VLBI (Very-Long-Baseline Interferometry) radio-synthesis telescope in the Southern Hemisphere; a new gamma-ray telescope at Woomera; and a revolutionary long-baseline optical interferometric telescope which will make possible for the first time the direct measurement of the diameters of thousands of stars.

A new National Science Centre, designed to be a "temple of science education", will open near the new parliament house. But Australians should take note that half of its A\$200 million cost was a bicentennial gift from the people of Japan.

Charles Morgan

God does not need science . . .

SIR—I should like to respond to some features of your article "Setback for creation science" (*Nature* 327, 643; 1987).

"Common sense" played no part in the US Supreme Court's decision against Louisiana's "Balanced Treatment Act", because "common sense" overwhelmingly supports the concept of an omnipotent creator over an ailing theory of evolution.

The "blend of fancy and whimsy" *Nature* refers to as creationist evidence was characterized by the two dissenting judges, Justice Antonin Scalia and Chief Justice William Rhenquist, as "the large body of uncontested evidence presented on behalf of creation science". As for the Supreme Court having "done the decent thing", it merely upheld one covert religious system (darwinism, or atheistic secular humanism) over an overt system (creationism).

The "uncertainties of the process of inference" *vis-à-vis* darwinism were recognized as soon as Darwin proposed his conjecture and even earlier when Alfred Russell Wallace preempted Darwin on darwinism. They are now being conceded belatedly by an embattled 128-year-old theory. I disagree that "contradictory evidence" against darwinism "is hard to find". We are drowning in it. Colin Patterson, senior palaeontologist at the British Museum, referred to molecular homology (a latter-day prop of darwinism) as "anti-knowledge" generating "anti-theory". He referred to so much "slop" in the data of molecular homology and pointed out that its processing had been "massaged with evolutionary theory".

Many of the founders of our science (for example Newton, Pasteur and Maxwell) believed in the same creator as the creation scientists *Nature* labelled as "charlatans". *Nature* errs seriously when it says that the conflict in reconciling evolution with religious beliefs is "not unavoidable". It is unavoidable. Religion predicates a creator, whereas darwinism was contrived to negate him. Creationism and darwinism constitute the opposing poles of the ultimate cosmic axis of God versus Satan, in the sense of the creator being the antithesis of the destroyer. This is theology, not science, but it underlies scientific view of origins.

On *Nature's* assertion that people who believe in *Genesis* are "gullible", it takes more gullibility to believe in darwinism than in *Genesis*. Darwin himself said that "to suppose that the eye . . . could have been formed by natural selection, seems, I freely confess, absurd in the highest degree".

Surely *Nature* does not seriously believe that people are "mischievous" when they decline to follow the criteria that other

people readily accept.

Creationism is an ancient idea intuitive to aeons of mankind. It is no more untenable a concept than the postulate of modern physics that matter can be created from nothing. Creationism may be a disgusting idea to many (for theological, political or personal reasons), but it has a clear right to a hearing on its merits as a plausible explanation of origins, on at least a par with darwinism. It becomes instantly implausible if, and only if, one is covertly (or otherwise) obliged to sustain atheism as a *sine qua non*. Finally, one bolsters such an *a priori* position by invoking a *deus ex machina* such as darwinism.

Incidentally, Darwin's only degree was in theology, not in science. Although he was "the naturalist" aboard the *Beagle*, he had no scientific training nor credentials. Yet, theologians who defend creationism are ridiculed by evolutionists for not being scientists.

God does not need science to prove His existence. In fact, in Christianity, the only prescribed way to the godhead is not by science, but by a little child's faith. Atheism, on the contrary, cannot sufficiently justify its existence on something as nebulous and negative as mere nonbelief in the creator. It desperately needs credentials, even pseudoscientific ones.

REGINALD T. CHELVAM

1740 West 27th Street, Suite 301,
Houston, Texas 77008, USA

A sideways look

SIR—A.C. Dornhorst (*Nature* 329, 758; 1987) cautiously suggests that El Greco suffered from unilateral astigmatism. He postulates that the artist viewed the subject with the abnormal (astigmatic) eye, but the canvas with the normal one. This is unlikely: severe unilateral astigmatism results in amblyopia — if he had used this eye to view the subject it would have been so defective he wouldn't have bothered to paint at all!

A.R. FIELDER
M.J. MOSELEY

University of Leicester,
Department of Ophthalmology,
Clinical Sciences Building,
Leicester Royal Infirmary,
PO Box 65,
Leicester LE2 7LX, UK

SIR—If El Greco's astigmatism were unilateral, the other eye being 'normal', he would doubtless have seen the world more clearly with the better, emmetropic eye, and would not have produced astigmatic pictures. If the fellow eye were highly myopic, however, the astigmatic eye having a spherical equivalent of approximately zero, he would have been obliged

to use the astigmatic eye to view his models, who would then appear distorted, while faithfully reproducing this perceptual distortion on the canvas using the other, spherically myopic eye. It is true that this would be an unusual state of refractive affairs, but then El Greco was an exceptional man.

It is not necessary to postulate this implausible form of anisometropia to explain his painting, however. If both eyes were astigmatic he would have developed meridional amblyopia¹ which would effectively blur lines in the axis of the more ametropic meridian, both in the world he viewed and in the one he represented on canvas, generating the predilection for lines and axes of a particular orientation, to the exclusion of their less clear perpendiculars, which distinguishes his work.

Philosophical nonsense it may have seemed, but the ophthalmologist's proposition had its origin in the viscous world of perceptual aesthetics, where logic follows the artist's private lights.

NICHOLAS EVANS

The Old Bakehouse,
Bruntingthorpe,
Lutterworth,
Leicester, UK

1. Gwiazda, J., Bauer, J., Thorn, F. & Held, R. *Clin. Vision Sci.* 1(2), 145–152 (1986).

Human genome sequence

SIR—The proposed project, in the United States, to sequence the entire human genome, has long-term implications that are disconcerting. This project would require the establishment of a large central bureaucratic organization to oversee and coordinate the effort, and such bureaucracies do not voluntarily dissolve themselves when the task for which they were originally set up has been accomplished. Instead, they find something else to do, much as NASA (National Aeronautics and Space Administration) proposed the ill-fated space shuttle, to justify its continued existence, once the task of 'beating the Russians to the Moon' had been accomplished. Thus the genome project, once initiated, would probably lead to a permanent drain on manpower and resources from other areas of biomedical research and a permanent tilt towards 'big science' of which high-energy physics is the best example. Anyone who knows how high-energy physics research is currently done, knows how much of a disaster that would be.

ROBERT J. YAES

Albert B. Chandler Medical Center,
Department of Radiation Medicine,
University of Kentucky,
Lexington, Kentucky 40536-0084,
USA

No prophecies for the superstitious

There is no way of telling what discoveries lie ahead, but some of the baggage of conundrums left unanswered in 1987 may be instructive.

SUPERSTITIOUS earlier generations used to mark new years with the publication of almanacs which, by judiciously blending near-certainty (the time of sunset) with speculation ("the King will die"), gave often-wide currency to unsubstantiable, but also irrefutable, soothsayings. In the nature of things, predicting the future course of discovery, the process of learning the unexpected, is a contradiction in terms. What follows, therefore, is not an almanac in the old sense, but a listing of some of the interesting questions left unanswered at the end of 1987.

The most clamant of these questions is also the most puzzling — that of knowing the mechanism of superconductivity in the ceramic 'high-temperature' superconductors. It is just 14 months since Müller and Bednorz announced the measurement of superconductivity at above 30 K in lanthanum-barium-copper oxide, and less than a year since Chu and his colleagues at the University of Houston described superconductivity at 90 K in an yttrium analogue.

Now it seems that there may be two classes of ceramic superconductors, with transition temperatures of 30 and 90 K, respectively. Certainly, those finding evidence for superconductivity nearer to room temperature have mostly been compelled, with disappointment, to account for their observations in terms of surface phenomena or as otherwise exceptional happenings (which are none the less intriguing on that account).

For the time being, much the same is true of the explanations of the new phenomena on offer. The standard theory of conventional superconductors is itself notoriously incapable of predicting transition temperatures accurately, but nobody should be surprised at that, given the complexity of all attempts to calculate phase transitions for cooperative systems.

Part of the difficulty with the new materials is that the interaction between electrons must entail exchange interactions between copper atoms through the intermediary of intervening oxygens (whence the suspicion that Anderson's resonating valence-bond theory must be on the right lines); the snag is the formidable complexity of integrating all this with the thermodynamics of a vibrating lattice, but there must surely be some break in the problem in the year ahead.

There can be few theoretical issues whose resolution would bring such

important prizes. While empirical methods have done so little to provide materials with higher transition temperatures, understanding seems the surest way forward. Meanwhile, there are many who will recall Professor V.L. Ginzburg's powerful case (at the *Nature* conference in November) that even liquid-nitrogen superconductors will be a great boon; letting the nitrogen boil off is a reasonable option.

The excitement over superconductivity is sometimes said to have been the best thing to have happened to solid-state physics for 20 years, but that is of course either a canard or an illusion. For quite different reasons, but in different ways, the field has been booming for several years, as shown by the now almost routine measurements of surface structure (with scanning tunnelling microscopes as well as diffraction techniques). What seems to be happening is rapid progress towards a state of affairs in which atomic structures are directly observable, by one technique or another. Can it be long before atom-by-atom fabrication is also possible?

Precision spectroscopy is similarly full of promise. Quite astonishing things are possible. In the latest issue of *Physical Review Letters*, for example, F. Diedrich *et al.* from the Max Planck Institute of quantum optics at Garching describe the measurement by spectroscopic techniques of the 'phase transition' between 'solid' and 'gaseous' clusters of laser-cooled magnesium ions held in an electromagnetic trap (there is a photograph of the fluorescent images of seven such ions held in an asymmetric hexagon); another West German group (J. Neukammer *et al.*, most from the Free University of Berlin) has measured barium atoms with an outer electron in a state whose principal quantum number exceeds 500 (which are sensitive measures of feeble electric fields); and F. de Martini *et al.* (from the University of Rome) have been able to demonstrate the prolongation of the lifetime of excited dye molecules in a Fabry-Perot optical cavity, thereby showing in the most direct fashion possible the coupling of atoms (or molecules, in this case) to the states of the radiation field in which they are immersed. Sooner rather than later, the precision of the new spectroscopy will yield new frequency standards, but there will be greater benefits along the way.

On the grander questions (What is matter? Where did it come from?) there is

at least some hope that the question of the 'fifth force' (the short-range accompaniment of gravity) will be settled one way or another during 1988. High-energy physicists appear content that the 'top' quark exists, although confirmation must probably await the next generation of accelerators. Meanwhile, a Japanese group working with the new KEK electron-collider seems to have gone further than others in ruling out a fourth generation of leptons (after electrons, muons and taus).

One setback for astronomers and for the rest of us is that the Hubble Space Telescope, potentially the most valuable of all Earth satellites so far, is still waiting on the ground in California for a now further delayed shuttle launch, while the Voyager spacecraft will not reach Neptune until next year.

Has molecular biology by comparison become flat, almost dull? That the question can even be asked about what has been the cutting edge of biology for the past 30 years points to what seems an important transformation of the field. That too is a canard, but questions such as the mechanism of the differentiation of cells, or tissues, seem almost as obscure now as a decade ago, chiefly because research has so far uncovered mostly the complexities.

In any case, the flood of discoveries in the past two years about the genetic basis of inherited diseases as different as muscular dystrophy and manic-depressive illness is certain to continue, and to yield an understanding of mechanisms in some cases. Indeed, that field, the unravelling of the intricacies of the immune system (with its bearing, among other things, on AIDS, acquired immune deficiency syndrome) and the demands of biotechnology are a reminder that, even after the huge growth of molecular biology in the past decade, there is now much more valuable work to do than there are skilled people to undertake it.

Meanwhile, it is natural but exciting that interest in neural networks — a way of modelling nervous systems — should have succeeded in capturing so much interest from a wide and interdisciplinary group of people, from molecular biologists to electrical engineers. Those insisting on some kind of a prediction for 1988 might take a chance of something in that field. But what?

John Maddox

Chemical physics

Energy flow in benzene

Marsha I. Lester

ADVANCES in high-resolution laser spectroscopy now permit a comprehensive examination of the flow of excess vibrational energy within large polyatomic molecules. New findings on vibrational energy exchange within the ground (S_0) and excited (S_1) electronic states of benzene were presented at a recent meeting*. Vibrational excitation can be localized in benzene C–H bonds, at lower internal energies than previously known, for as long as a picosecond. The excess energy then redistributes over a high density of nearly isoenergetic vibrational background states. The coupling to optically inactive background states changes with the rotational quantum numbers of the prepared vibrational state. Further, specific chemical functional groups enhance the rate of energy randomization.

The vibrational motions of large molecules (with N atoms) at low energies can be described by $3N - 6$ independent vibrational degrees of freedom — the normal modes of the whole molecule. As the vibrational energy content of the molecule is increased, anharmonic couplings between normal modes lead to the eventual breakdown of this description. A different framework emerges, in which a few vibrational motions decouple from the dense background of vibrational states. These vibrations are usually localized on molecular bonds such as the high-frequency C–H, O–H or N–H stretches. In recent experiments (Y.T. Lee, University of California, Berkeley), the C–H stretching mode of S_0 benzene was prepared with an increasing number, ν , of vibrational quanta to look for the onset of local mode character in the vibrational motion¹.

The fundamental and overtones of the C–H stretching mode of benzene exhibit broad spectral features at room temperature², attributed to spectral congestion and/or lifetime (homogeneous) broadening. To examine the homogeneous broadening of individual rotational lines, benzene is first cooled to 5 K in a supersonic jet expansion, which reduces the breadth of the envelope of unresolved rotational lines to 4 cm^{-1} . A state-selective labelling technique is used to resolve the C–H overtone features (Fig. 1). The population in selected rotational levels in the vibrational ground state is monitored by resonant two-photon ionization. As a tunable infrared laser pumps the C–H vibrational levels, the vibrational ground-

state population becomes depleted and the spectrum is revealed by the decreased ion signal.

The resolved C–H overtone spectra (Fig. 2) are dramatically different from room-temperature spectra. The fundamental region ($\nu=1$) exhibits three peaks because of 'Fermi' resonances between the C–H stretch and combination states of nearly the same energy. At $\nu=2$, more than 20 peaks appear over a 200-cm^{-1} spectral range. The number of peaks increases because of the higher density of normal-mode combination states to which the C–H stretch couples. The width of each peak in the $\nu=1$ and 2 spectra is determined by the residual rotational envelope. At $\nu=3$, however, only four peaks are evident in the overtone spectrum, occupying a substantially narrower spectral range than $\nu=2$. The narrower frequency range signals the onset of local-mode behaviour for the C–H stretch at $\nu=3$. The C–H stretch becomes localized because of less efficient coupling with background states.

The primary feature in the $\nu=3$ overtone spectrum is 10 cm^{-1} wide, which is taken as an upper limit for the homogeneous line width in benzene C–H overtone spectra. This indicates (by application of Heisenberg's uncertainty principle) that the local-mode character of the C–H

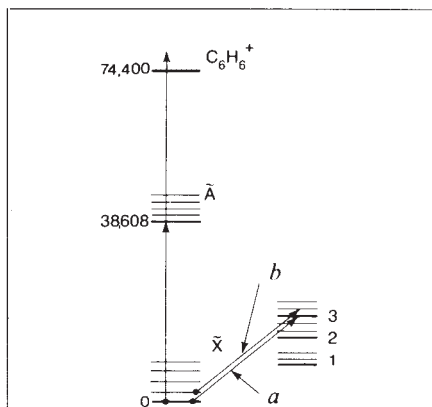


Fig. 1 The method of Y.T. Lee and co-workers for observing the overtone vibrational spectrum of benzene. Two photons from an ultraviolet laser photoionize ground-state molecules. A second tunable infrared laser excites the C–H vibrations (a), depleting the ground-state population and hence reducing the ion signal. Because resonant two-photon ionization is state selective, 'hot-band transitions' (b) originating from low-frequency vibrational levels are avoided. Energies in cm^{-1} indicated to the left. ν for the local modes indicated to the right. (From ref. 1.)

stretch in benzene persists for more than 0.5 ps. The vibrational lifetime is a measure of how long energy is confined in the C–H bond before decaying into a dense manifold of background vibrational states. This lifetime is substantially longer than reported in previous experiments and theoretical investigations of higher C–H overtones.

The S_1 benzene, an abrupt diffuseness in the absorption spectrum and a dramatic drop in the fluorescence quantum yield (the ratio of emitted to absorbed photons) occurs at an excess vibrational energy of $3,000\text{ cm}^{-1}$, indicating the opening of a non-radiative decay channel in benzene. The origin of this process has long been a mystery. High-resolution sub-Doppler spectroscopy has recently been used to examine individual rotational levels in this energy regime (E.W. Schlag, University of Munich). Through this work, the spectral changes in this energy region are now attributed^{3–5} to the onset of intramolecular redistribution (IVR) within S_1 .

Doppler-free excitation of benzene is achieved by the simultaneous absorption of two photons from equal and counter-propagating travelling light waves. The fluorescence excitation spectrum induced by Doppler-free two-photon excitation is recorded as the excess vibrational energy in the S_1 manifold is increased. The band at $3,412\text{ cm}^{-1}$ has few rotational lines compared with the rich spectra found at lower excess energies. All lines with rotational quantum number $K' > 0$ (K' is the angular momentum on the six-fold symmetry axis) are absent³. Furthermore, the lifetime of the $K'=0$ levels decreases as the total rotational angular momentum J' is increased, as shown in both frequency⁴ and time-domain⁵ experiments.

The rotational-level dependence is explained by a primary IVR process within S_1 , arising from Coriolis coupling of the prepared state with optically inactive background levels. Background levels having out-of-plane vibrational character are short-lived and lifetime broadened by rapid nonradiative relaxation. The non-radiative decay is presumably internal conversion from S_1 to S_0 , which is independent of rotational quantum state. The onset of the dynamic IVR process is responsible for the sudden increase in the nonradiative relaxation rate and the corresponding break-off of fluorescence which characterizes the $3,000\text{ cm}^{-1}$ excess-energy regime of S_1 benzene.

The above cases exemplify vibrational levels that are relatively uncoupled from other molecular motions so that energy redistribution is slow. A chemical functional group has been identified (C.S. Parmenter, Indiana University) which has the opposite effect, accelerating IVR. Attaching a methyl (CH_3) group to an aromatic-ring system substantially

*International Laser Science III Meeting, Atlantic City, 1–5 November 1987.

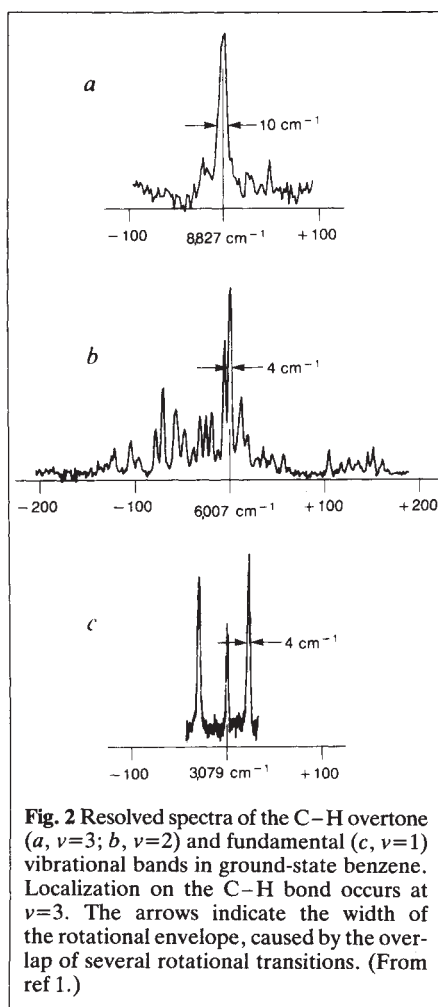


Fig. 2 Resolved spectra of the C-H overtone (*a*, $\nu=3$; *b*, $\nu=2$) and fundamental (*c*, $\nu=1$) vibrational bands in ground-state benzene. Localization on the C-H bond occurs at $\nu=3$. The arrows indicate the width of the rotational envelope, caused by the overlap of several rotational transitions. (From ref 1.)

enhances the rate of IVR (ref. 6). The nearly-free CH_3 rotor increases the density of vibrational-rotor states at a given excess energy and induces effective mixing of vibrational states. The level mixing in *p*-fluorotoluene is evaluated from the spectral congestion in the dispersed S_1 - S_0 fluorescence and by gating the S_1 emission via chemical timing experiments. Repulsive van der Waals interactions between the hydrogens of the methyl group and adjacent atoms in the aromatic ring are shown to be the source of vibrational level mixing⁷. The symmetry of the hindered rotor imposes selection rules for the coupling of vibrations to rotor states which are consistent with spectroscopic observations. □

1. Page, R.H., Shen, Y.R. & Lee, Y.T. *Phys. Rev. Lett.* **59**, 1293-1296 (1987).
2. Reddy, K.V., Heller, D.F. & Berry, M.J. *J. chem. Phys.* **76**, 2814-2837 (1982).
3. Riedle, E., Neusser, H.J. & Schlag, E.W. *J. phys. Chem.* **86**, 4847-4850 (1982).
4. Riedle, E. & Neusser, H.J. *J. chem. Phys.* **80**, 4686-4692 (1984).
5. Schubert, U., Riedle, E., Neusser, H.J. & Schlag, E.W. *J. chem. Phys.* **84**, 6182-6189 (1986).
6. Parmenter, C.S. & Stone, B.M. *J. chem. Phys.* **84**, 4710-4711 (1986).
7. Moss, D.B., Parmenter, C.S. & Ewing, G.E. *J. chem. Phys.* **86**, 51-61 (1987).

Marsha I. Lester is in the Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

Medical genetics

The slow road to gene therapy

D. J. Weatherall

WHEN thoughts first turned to replacing defective human genes, the most obvious candidates were the genetic disorders of haemoglobin. After all, more was known about their molecular pathology than any other genetic diseases and they are the commonest inherited diseases of man. But it soon became apparent that there would be enormous difficulties in working with globin genes. A paper by Dzierzak and her colleagues on page 35 of this issue¹, however, suggests that the gene-therapy field may have gone full circle and that globin is again a prime candidate.

Red blood cells, together with granulocytes, macrophages and platelets, are the terminal differentiation products of a self-renewing haemopoietic stem-cell population. To transfer 'foreign' globin genes to cure a disease like thalassaemia, which results from defective α - or β -globin chain production, several problems would have to be overcome. First, it would be necessary to transfect stem cells, which make up only a small proportion of the bone marrow population and many of which appear to be out of cycle at any one time. Second, it might be necessary to provide a mechanism for preferential proliferative advantage of the transfected stem cell population. And, most formidably, it would be essential to ensure that the transfected globin genes were expressed only in red cell precursors and that they are tightly regulated so as not to cause imbalanced globin chain production; the conversion of α - to β -thalassaemia, or vice versa, although a *tour de force* for molecular biologists, would not be so exciting for the unfortunate patient.

Because none of these problems seemed likely to be solved easily, many workers turned their attention from globin disorders to some of the inborn errors of metabolism resulting from deficiency of 'housekeeping' genes, that are expressed in many cells and which do not appear to require such tight regulation as the globin genes. Two conditions, the Lesch-Nyhan syndrome, caused by deficiency of hypoxanthine phosphoribosyl transferase (HPRT), and severe combined immune deficiency caused by defective production of adenosine deaminase (ADA), became the prime candidates. An enormous amount of work has been carried out to develop methods for transfer of these genes; it has been suggested, perhaps irreverently, that there may be more workers trying to correct ADA deficiency than there are patients with the disease. But despite all this activity

progress has been slow, largely because of initial difficulties in obtaining an adequate efficiency of transfection and, in particular, in achieving useful levels of expression of the transfected genes.

Various approaches to gene transfer have been investigated, including direct insertion by the uptake of calcium micro-precipitates of DNA², microinjection² or electroporation³, the use of retrovirus vectors⁴⁻¹², and targeted modification by homologous recombination¹³. Currently, the retrovirus field is the most active, but although progress has been made both in cell-culture models and *in vivo*, and many of the difficulties of inefficient transfection have been overcome, there remain the problems of adequate levels of expression and the continued stability of transferred genes after many rounds of cell division.

Unlike many previous studies, which have involved retrovirus vectors containing human complementary DNA sequences driven by various viral transcriptional signals in the long terminal repeat sequences or other viral control regions, Dzierzak *et al.* have inserted intact human β -globin genes into several different constructs¹. It appears that, when stem cells transfected with these genes are injected into lethally irradiated mice, the human globin genes are expressed exclusively in erythroid cells of the recipient for periods ranging between four and nine months. And this occurs whether or not viral transcriptional control signals are present. These results suggest that sequences required for tissue specificity of expression are included in the transfected human DNA sequences.

Although this result is very encouraging, the problem of low levels of expression remains. But there is light on the horizon. Recent studies by Grosfeld and his colleagues¹⁴ have identified sequences located 50 kilobases upstream and 20 kilobases downstream of the human β -globin gene which, when juxtaposed to β -globin genes, dramatically increase their expression in transgenic mice containing a single copy of the sequence. The high level of expression appears to be independent of the chromosomal site of insertion and is also tissue-specific. It should now be possible to incorporate these sequences in retroviral constructs of the type used by Dzierzak *et al.*¹ in the hope that tissue-specific expression of relatively large amounts of human globin can be achieved. Whether equivalent sequences exist for the housekeeping genes mentioned above remains to be determined.

Although these results¹⁻⁴ represent a genuine advance towards gene therapy, many difficulties remain. In the case of the globin disorders, sufficient stem cells will have to be transacted to provide progeny capable of repopulating adequate numbers of haemopoietic progenitors. Because only a small proportion of stem cells contributes to the differentiation compartments of the marrow at any one time this may be particularly difficult. And because the beneficial effect of inserted globin genes is only manifest in these compartments, that is, in red cell precursors, it may still be necessary to provide some proliferative selective advantage to the transfected stem cells. And even if these problems are overcome, extensive experiments on animals will be required to determine the safety of the use of retrovirus vectors. It is vital to ensure that the vectors do not become a cancer hazard through recombination events, for example, or as a result of random insertion into a large number of cells by the activation of oncogenes. And it will be most important to ensure that expression of induced genes is stable over a long period.

At a time when there is much concern about human genetic manipulation, it is important to emphasize that all these experiments are connected with somatic gene therapy — that is, the removal of human cells, transfection of appropriate genes and their subsequent replacement. Despite the remarkable successes of transgenic experiments in mice and other

species, in my opinion this approach should not be applied to humans as it would alter the genetic make-up of subsequent generations. It is for this reason, and to emphasize the importance of adequate animal work to ensure the safety of gene-transfer technology, that guidelines have been drawn up by the National Institutes of Health and by the European Medical Research Councils, which provide sensible rules about the measures to be taken before the first attempts at human gene replacement. Although there is cause for cautious optimism, it is likely that it will still be some time before we are ready to replace defective human genes. □

1. Dzierzak, E.A., Papayannopoulou, T. & Mulligan, R.C. *Nature* **331**, 35–40 (1988).
2. Anderson, W.F. *Science* **226**, 401–409 (1984).
3. Chu, G., Hayakawa, H. & Berg, F. *Nucleic Acids Res.* **15**, 1311–1326 (1987).
4. Williams, D.A. & Orkin, S.H. *J. clin. Invest.* **77**, 1053–1056 (1986).
5. Anderson, W.F. *et al. Cold Spring Harb. Symp. quant. Biol.* **51**, 1065–1072 (1986).
6. Miller, A.D., Palmer, T.D. & Hock, R.A. *Cold Spring Harb. Symp. quant. Biol.* **51**, 1013–1020 (1987).
7. Dick, J.E. *et al. Cell* **42**, 71–79 (1985).
8. Keller, G. *et al. Nature* **318**, 149–154 (1985).
9. Lemischka, I.R., Raulet, D.H. & Mulligan, E.C. *Cell* **45**, 917–927 (1986).
10. Williams, D.A., Orkin, S.H. & Mulligan, R.C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2566–2570 (1986).
11. Joyner, A. *et al. Nature* **303**, 556–558 (1983).
12. Williams, D.A. *et al. Nature* **310**, 476–480 (1984).
13. Gregg, R.G. & Smithies, O. *Cold Spring Harb. Symp. quant. Biol.* **51**, 1093–1100 (1987).
14. Grosveld, F. *et al. Cell* (in the press).

D. J. Weatherall is Nuffield Professor of Clinical Medicine and Director of the MRC Molecular Haematology Unit, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

Cosmology

A nearly primordial galaxy

Wallace L. W. Sargent

DIFFERENT types of galaxy are thought to differ mainly in the rate at which they consume their primordial gas to make stars and thence to produce heavy elements. Elliptical galaxies probably use up most of their initial gas during their initial collapse — probably in the first billion (10^9) years. At the other extreme, the irregular galaxies (which are practically always dwarf objects) maintain a more-or-less steady average rate of star formation, perhaps occurring in the form of punctuated bursts of activity, for their whole life spans of around 10 billion years. The spirals are intermediate between these two extremes: the spheroidal central bulge forms in an early-collapse phase; while a substantial fraction of their gas settles into a slowly rotating disk which is consumed more slowly. It is thus expected that most galaxies, particularly the ellipticals, were much more active and therefore brighter in the remote past. Locally, there is little sign of recent galaxy formation. The few candidates for young

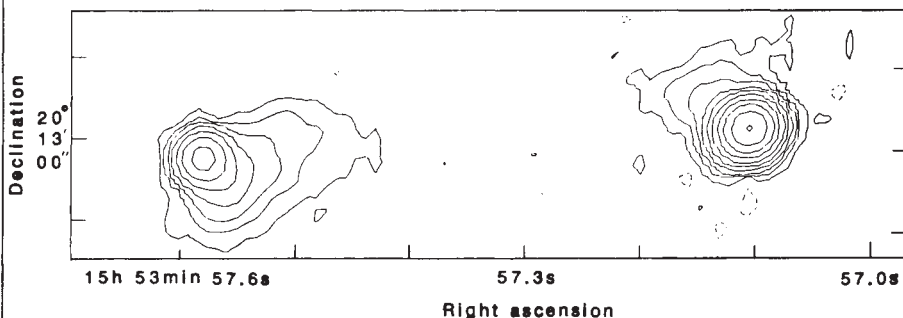
galaxies are practically all dwarf systems and past low-level activity could be obscured by a current burst of star formation. Hyron Spinrad and colleagues, however, now report (McCarthy, P.J. *et al. Astrophys. J.* **319**, L39–L44; 1987) the observation of a distant galaxy which, even if not primordial, is in a very early stage of development.

It has always been supposed that a primordial galaxy would be most easily

identified through its Lyman- α emission line. Lyman- α is radiated copiously by the recombination of electrons with hydrogen ions in the hot, ionized gas around young, hot stars. Moreover, Lyman- α should still be observed in a proto-galaxy which has not yet had time to transform hydrogen into heavier elements. The Lyman- α line, emitted at a wavelength of 1,216 Å, can only be observed from the ground in the spectra of objects with a redshift greater than about $z = 1.6$. This is one reason why the search for proto-galaxies has been so far unsuccessful.

During the past few years, Spinrad and his colleagues at the University of California at Berkeley have had some success in locating associated objects near quasars or radio galaxies of known redshift by imaging the surrounding field through a narrowband filter tuned to the Lyman- α line. This picture is then compared with one obtained through a broadband filter centred on the same wavelength. Objects that emit most of their radiation in the Lyman- α emission line stand out even if they are near the limit of the pictures. They used this method, for example, to find a very faint galaxy at a redshift of $z = 3.215$ close to the quasar PKS1614+051 (Djorgovski, S. *et al. Astr. J.* **93**, 1318–1325; 1987). This galaxy is by far the most distant galaxy known, but its spectrum also contains emission lines of elements other than hydrogen and it is not a proto-galaxy.

Now, McCarthy *et al.* report the discovery of a large cloud of ionized gas with a redshift of $z = 1.82$, associated with the radio galaxy 3C326.1, that has many of the characteristics of a primordial system. Maps obtained with the Very Large Array radio telescope reveal the radio source to be a small double with the components separated by only 7 arc s (see figure). The Lyman- α image reveals a cloud of ionized hydrogen that encompasses the radio sources; the diameter of the cloud is about 100 kpc. The broadband images show two very faint, blue stellar objects at the edge of the cloud as well as some extremely blue, diffuse objects near its brightest parts. Spectra indicate that the blue objects are likely to be continuum emission from hot stars. The spectra show that, as expected, the cloud radiates strongly in



Total intensity radio map of 3C326.1 by McCarthy *et al.* showing the small double.

the Lyman- α emission line, which is Doppler broadened by speeds up to about 1,000 km s⁻¹. A spectrum with the slit aligned along the long axis of the cloud gives some indication of a systematic velocity gradient of about ± 500 km s⁻¹. There are also weak emission lines from ionized carbon, radiated at wave-lengths of 1,909 Å (C III) and 2,325 Å (C II), with the same redshift as the Lyman- α and similarly distributed over the face of the cloud.

Hence the cloud is not a primordial object because it already contains heavy elements. D. Kunth and I, however, in a discussion on the absence of nearby primordial galaxies (*Astrophys. J.* **300**, 496–499; 1987) showed that the hot stars which ionize the gas also produce and eject heavy elements on a timescale of only a few million years. Therefore, it is quite possible that the ionized (invisible) parts of the cloud are pure hydrogen and helium formed in the Big Bang.

The source of the ionizing radiation in the cloud is indeed likely to be young, hot, massive stars, as there is no indication either in the optical or the radio data of quasar-like activity that could provide an alternative source of ionizing photons. As McCarthy *et al.* point out, the origin of the strong radio emission is mysterious. It is possible, however, that the burst of star formation which is inferred may have been triggered by the motion of the radio blobs through the gas; similar phenomena have been observed recently in nearby radio galaxies. The inferred rate of star formation is about 70 solar masses ($M_{\odot}$) per year in the form of stars more massive than $10 M_{\odot}$ and about $500 M_{\odot}$ per year if a correction is made for the mass in stars that produce negligible ionizing radiation.

McCarthy *et al.* estimate that the mass of ionized hydrogen in the cloud is less than $10^{11} M_{\odot}$ on the assumption that the gas is uniformly distributed throughout its volume. This upper limit could be reduced by orders of magnitude if the gas is clumped. Thus, if it is distributed in dense regions of ionized hydrogen around hot stars, only about $10^7 M_{\odot}$ would be required. The authors remark that the total mass in stars and gas is unknown. The line width and possible velocity gradient, however, imply a very large total mass, about $2.8 \times 10^{12} M_{\odot}$, if the cloud is gravitationally bound; they do not address this question. The cloud of ionized gas discovered by McCarthy *et al.* is the nearest approach to a primordial galaxy yet discovered. As they point out, if the cloud really is a proto-galaxy it has made a powerful non-thermal radio source *before* forming most of its stars. Also, the object is not shrouded in dust as some authors suggest primordial galaxies might be. □

Wallace L.W. Sargent is at the Palomar Observatory, California Institute of Technology, Pasadena, California 91125, USA.

AIDS

Receptor molecule blocks HIV

Robin A. Weiss

SINCE the CD4 (T4) antigen was identified^{1,2} as an essential component of the cell surface receptor for the human immunodeficiency virus (HIV-1), it has become one of the most intensively studied receptors of any enveloped virus. Now five papers, one in *Science*³, and four in this issue of *Nature*^{4–7} show that soluble forms of CD4 antigen (sCD4) can block the infectivity of the virus. Clinical trials of sCD4 as an antiviral agent are already being planned in the United States.

CD4 antigen acts as the HIV-1 receptor through the recognition by the outer viral envelope antigen, gp120, of a site on the receptor molecule⁸ possessing an affinity constant⁹ of about 10^{-9} M. Transfer and expression of the CD4 gene to human cells such as HeLa confers susceptibility to HIV-1 infection¹⁰. (Curiously, this is not the case for mouse cells; although HIV-1 specifically binds to mouse cells expressing human CD4, it does not infect them.) Monoclonal antibodies against CD4 block the binding of HIV-1 and subsequent infection and syncytium induction^{1,2,8,11}. Thus, the CD4 antigen presents the initial binding site for HIV-1 to enter the host cell.

The reports of sCD4 indicate that saturating amounts are needed to block HIV-1 infectivity. In most of the studies complete inhibition of low-titre HIV-1 (100–1,000 tissue-culture infectious doses, TCID) was achieved only with microgram amounts of sCD4. Although the Genentech group³ shows very high affinity of its sCD4 with gp120, it was barely able to achieve complete blocking of HIV-1 infectivity (100 TCID) with undiluted sCD4 preparations. In contrast, the Smith Kline & French group⁶ observes greater than 3,000-fold inhibition of HIV-1 infectivity with $8.6 \mu\text{g m}^{-1}$ sCD4. The inhibitory effect of the other sCD4 molecules is intermediate, whether produced from insect⁵ or mammalian^{6,7} cells. This could well reflect variation in the HIV-1 infectivity assays rather than the efficacy of different sCD4 preparations.

The precise binding site on CD4 for viral gp120 is now being analysed in several laboratories by site-specific mutagenesis. Several lines of study indicate that it is near to the epitope recognized by leu 3a and related monoclonal antibodies. These strongly inhibit HIV-1 binding, infection and syncytium induction¹¹, and some anti-idiotypic sera to leu 3a weakly react in virus neutralization assays^{12–14}. Moreover, our recent data indicate¹⁵ that the epitope recognized by leu 3a is highly conserved in primate

evolution, and that cell culture infection by HIV-1 of New World monkeys such as marmosets is blocked by the monoclonal antibody leu 3a. One would therefore surmise that HIV-1 binds to a CD4 epitope necessarily conserved for the normal physiological role this antigen plays in regulation of the immune system, thought to be interaction with class II antigens of the major histocompatibility complex¹⁶. If so, one might expect sCD4 itself to be immunosuppressive. However, Reinherz's group shows⁵ that its sCD4 does not inhibit class II-dependent immune killing by a CD4⁺ cytolytic cell clone. This result provides encouragement for clinical trials of sCD4, which may serve *in vivo* not only to block spreading HIV-1 infection, but also to sequester free viral gp120 and reduce its possible immune toxicity in AIDS (acquired immune deficiency syndrome).

It remains to be determined how small a fragment of CD4 is able to inhibit HIV-1 infectivity. Most of the sCD4 preparations reported^{3–7} comprise almost all of the extracellular domains of the antigen, although Trauneker *et al.* show⁷ that a fragment containing only the two amino-terminal immunoglobulin-like domains is almost as inhibitory as the whole extracellular portion. All the groups^{3–7} appear to have used a single strain of HIV-1 in the sCD4 inhibition tests, but our own unpublished observations indicate that sCD4 inhibits all strains of HIV-1, HIV-2 and related simian immunodeficiency viruses tested. This is not surprising as all these viruses appear to recognize the same domain of CD4 as their binding receptor¹⁴.

HIV entry to non-lymphoid cells might be mediated by receptors other than CD4, including the infection of monocytes via Fc receptors. It is important to determine whether sCD4 inhibition of HIV infection applies solely to CD4-mediated entry. □

1. Dalglish, A.G. *et al.* *Nature* **312**, 736–766 (1984).
2. Klatzmann, D. *et al.* *Nature* **312**, 767–768 (1984).
3. Smith, D.H. *et al.* *Science* **238**, 1704–1707 (1987).
4. Fisher, R.A. *et al.* *Nature* **331**, 76–78 (1988).
5. Hussey, R.E. *et al.* *Nature* **331**, 78–81 (1988).
6. Deen, K.C. *et al.* *Nature* **331**, 82–84 (1988).
7. Trauneker, A., Lüke, W. & Karjalainen, K. *Nature* **331**, 84–86 (1988).
8. McDougal, J. *et al.* *Science* **231**, 382–385 (1986).
9. Lasky, L. *et al.* *Cell* **50**, 975–985 (1987).
10. Maddon, P.J. *et al.* *Cell* **42**, 685–698 (1986).
11. Sattenau, O.J., Dalglish, A.G., Weiss, R.A. & Beverley, P.C.L. *Science* **234**, 1120–1123 (1986).
12. Chanh, T. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 3891–3895 (1987).
13. Dalglish, A.G. *et al.* *Lancet* **ii**, 1047–1050 (1987).
14. Sattenau, O.J. *et al.* *Int. Conf. AIDS III* **160** (1987).
15. McClure, M.O. *et al.* *Nature* **330**, 487–489 (1987).
16. Doyle, C. & Strominger, J.L. *Nature* **330**, 256–259 (1987).

Robin A. Weiss is at the Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB, UK.

Quantum optics

Two-photon micromasers

Peter Knight

NORMAL lasers and masers emit numerous photons by stimulated emission from a macroscopically large number of excited atoms. The excitation which pumps the atoms into their excited states has to be particularly strong because an inversion in the usual thermodynamic equilibrium populations must be produced at a rate commensurate with the rate of loss by the uncontrollable decay of excitation. Indeed, lasers have become synonymous with intense coherent light. It might seem perverse therefore for quantum opticians to expend great effort producing devices that involve only a handful of atoms and stimulated photons. These so-called micromasers (see my News and Views article in *Nature* 326, 329; 1987) in fact hold the key to understanding peculiar quantum features of light amplification that are swamped in their high-intensity big brothers. A group in Paris, led by Serge Haroche, now reports (Brune, M. *et al.* *Phys. Rev. Lett.* 59, 1899–1902; 1987) a micromaser based on the simultaneous emission of two photons that is used to look at these quantum effects.

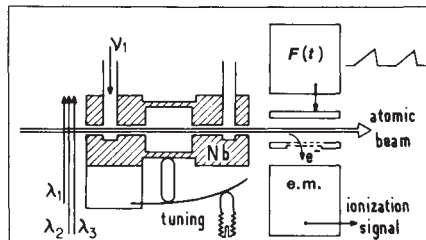
Quantum fluctuations in the emitted light intensity are usually small and involve few spontaneously emitted photons compared with 10^{10} stimulated photons in a normal continuous laser beam. A micromaser involves perhaps 30 photons with only five or ten atoms at a time in the emission process provided by a low-flux atomic beam. The atoms are excited into very high-lying 'Rydberg' states and persuaded to emit microwave radiation in a very high-quality superconducting cavity.

It is hard to detect low-energy microwave photons, especially when the power radiated by the atoms is only 10^{-18} W. Instead, the micromaser emission is detected indirectly by its effect on the state of the atoms involved which are interrogated by a technique called field ionization. Atoms that have made the transition to the lower micromaser state are counted as they leave the superconducting cavity and their numbers suddenly increase when maser action starts. The atoms are pumped into the excited state by a combination of lasers and occasionally microwave beams just before they enter the maser cavity.

Using the microwave technique, Haroche and co-workers have now built the first quantum oscillator working on two-photon stimulated emission where each atom emits two photons. The transition chosen is one that is not possible by single-photon emission. The two-photon maser has remarkable quantum

characteristics: it is triggered by the vacuum fluctuations of the quantized cavity field. Long start-up delays of several seconds before emission can be observed close to the maser threshold, where emission just balances the losses.

Haroche and co-workers use a beam of rubidium atoms excited into the $40S_{1/2}$ level, emitting pairs of photons with a frequency 68.41587 GHz which matches the maser cavity frequency. (Actually, it didn't quite, so a deformation applied with a screw bent the cavity and corrected the mismatch.) The transition is a genuine



The experimental scheme of Brune *et al.* The rubidium atom beam is first excited by three different lasers (λ_{1-3}) to a level above the $40S_{1/2}$ starting point of the two-photon transition. A microwave transition down to the $40S_{1/2}$ level is stimulated by the radiation ν_1 . The two-photon transition occurs in the cavity in the superconducting niobium (Nb), which has to be tuned by distortion. Atoms are field ionized in the ramped electric field $F(t)$ and freed electrons detected by the electron multiplier (e.m.).

two-photon transition and not two separate one-photon transitions in a cascade. The superconducting cavity was made at the European Nuclear Research Organisation CERN (a good example of the benefits to small science of major international laboratories) and has a quality factor of 10^8 so that radiation in the cavity was stored for 0.2 ms.

Haroche and his group operated their maser with only about 30 photons stored in the cavity to see the interesting low-intensity quantum effects. The flux of excited atoms necessary to maintain the radiation against the cavity loss is so low that only five atoms need be in the cavity at any one time. This is because the Rydberg atoms have enormous dipole moments which lead to a very strong two-photon coupling of atoms to field. If the flux of excited atoms is increased until the threshold is reached, the maser turns on, but only after a delay, sometimes of seconds. This delay is the most interesting feature of the two-photon micromaser. It occurs because the emission is initiated by quantum fluctuations in the cavity and not by a spontaneously emitted photon. A pencil balanced on its point takes a while to fall over and does so only because of random noise fluctuations. Similarly, the unstable excited atoms in the micromaser cavity are metastable and need a kick from the randomly fluctuating quantized radiation before parting with their excitation. What is surprising about the micromaser is that macroscopically long time delays are necessary near the maser threshold before the quantum emission starts. □

Peter Knight is in the Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, UK.

Palaeoecology

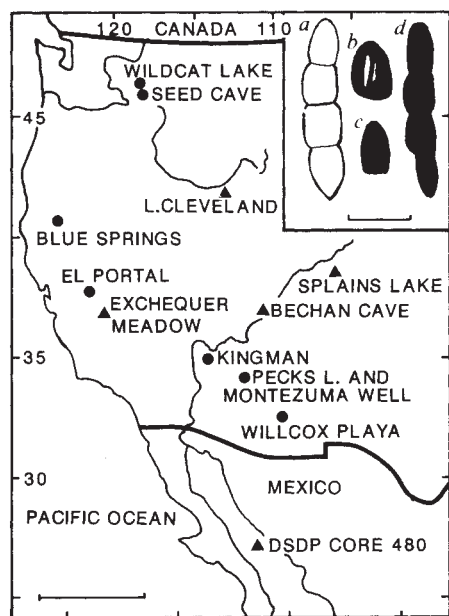
Causes of megafaunal extinction

Peter D. Moore

COCK ROBIN has nothing on the Quaternary megafauna when it comes to murder mysteries. The debate still rages concerning the relative roles of climatic change and human activities in the rash of extinctions of large herbivores and carnivores during the early part of the current interglacial, and palaeoecologists face several problems in trying to assemble a complete picture. First, how can we detect the loss of a large animal from a palaeoecosystem other than by a decrease in fossil frequency? The direct use of fossils is blighted by their relative rarity, so perhaps an indirect approach can be used — looking for megafaunal dung, for example, or even for the organisms that inhabit their dung! Second, how can we detect the synchronicity among species extinctions that might be expected from a

human cause? Further well-dated studies are required to provide this information². Third, how can we detect the human presence and evaluate the environmental impact of human cultures? This calls for various techniques, both archaeological and palaeoecological, such as the reconstruction of contemporaneous vegetation from the evidence of fossil pollen and the frequency of fires from charcoal fragments³.

The dung of the megafauna has been studied extensively, both to provide evidence relating to the diet of these animals and as a means of dating their period of existence. For example, mammoth dung was found⁴ in a cave in Colorado, dating from about 14,000 to 11,000 years ago, and dung from the Shasta ground sloth in the south-west United States⁵ became



The western United States and Mexico showing where *Sporormiella* spores were found in ●, historic and ▲, late-glacial sediments. Scale bar, 450 km. Top right: a, *Sporormiella intermedia* spore; b, fossil spores from mammoth dung, Bechan Cave; c, fossil spores from DSDP core 480; d, modern *Sporormiella minima* spores. Scale bar, 10 μ m. (Modified from ref. 1.)

considerably less abundant after about 11,000 years ago, corresponding to the time when Clovis hunting communities were active there. Davis¹ has now gone one step further than the dung hunters and has studied the dung-inhabiting fungus *Sporormiella*, whose spores he found preserved in the sediments of lakes in the western United States (see figure). These spores are abundant only in lakes around which there is a high local density of large herbivores. In fossil sediments they became very prominent following the introduction in historic times of domesticated herbivores, but Davis also finds them to be frequent in the sediments of late-glacial times, before 11,000 years before present (BP). They decline markedly at about this date, providing further evidence about the date of megafaunal extinction. So, even in the absence of skeletal evidence, the decline of the megafauna can be observed and dated.

Dating, of course, is critical in the debate concerning the causes of megafaunal extinction. Grayson has examined² the available dates and finds some important gaps in our knowledge of the time periods during which some members of the megafauna were present in North America. Only 7 of the 35 extinct genera considered had firm dates from the terminal Wisconsin (12,000–10,000 BP). Many of the megafauna, Grayson contends, may not have survived until 12,000 years ago, thus complicating any explanation of the mechanisms involved in their extinction.

If the human predation was the cause of

extinction, one would expect a correlation between archaeological and extinction dates, which seems to occur in the case of the Shasta ground sloth. But even when such correlation exists, it still provides only circumstantial evidence and does not demonstrate a causative relationship. Further information concerning the contemporaneous environment could be useful, especially if signs of human activity can be detected over wide areas. Take the case of Madagascar, for example. Seventeen megafaunal species disappeared from this island in relatively recent times, in the latter part of the Holocene (the past 10,000 years). The association between such extinctions and human settlement (dating around 2,000–1,500 BP) is compelling, and explanations of the precise nature of these events have often been linked with the increasing frequency of fires and consequent habitat changes on the island.

Burney has examined³ this hypothesis by analysing the charcoal content of lake cores from the centre of Madagascar. He finds that in the past 10,000 years the highest levels of charcoal abundance occurred between approximately 10,000 and 4,000 BP. Charcoal then declined very strongly and remained low for more than 2,000 years, only rising steeply at about 1,200 years ago. From these data he

concludes that fire was important in the ecology of Madagascar through the early part of the Holocene. Although fire had become less frequent and/or extensive before human occupation of the island, it is unlikely that the animals and plants of Madagascar would have been particularly fire-sensitive. It is possible, of course, that the human-induced fires of more recent times differ from the old fires in their timing, duration, extent or intensity, but their overall impact on lake charcoal influx is considerably less than the fires of the early Holocene. The simple explanation of megafaunal extinction as a result of habitat modification by human-induced fires is not, therefore, supported by the charcoal data. Direct predation, of course, is still a possibility.

So, if human agencies were involved in the death of the megafauna, they covered their tracks pretty well. Proving human involvement is likely to be an arduous task, but establishing a climatic link is undoubtedly even more difficult. □

1. Davis, O.K. *Quat. Res.* **28**, 290–294 (1987).
2. Grayson, D.K. *Quat. Res.* **28**, 281–289 (1987).
3. Burney, D.A. *Quat. Res.* **28**, 274–280 (1987).
4. Mead, J.L., Agenbroad, L.D., Davis, O.K. & Martin, P.S. *Quat. Res.* **25**, 121–127 (1986).
5. Lewin, R. *Science* **221**, 1036–1037 (1983).

Peter D. Moore is Reader in Ecology at King's College London, Campden Hill Road, London W8 7AH, UK.

Evolution of behaviour

Burrowing by vertebrates

M.J. Benton

TODAY, many vertebrates form burrows for living, feeding or protection — familiar examples include rabbits, moles, certain snakes, lizards and amphibians, and lungfish. The burrows can be well-formed semi-permanent underground systems, often of considerable complexity, but most are temporary passages forced into surface layers of the soil. The evolution of burrowing habits can be studied by analysing the behaviour and ecology of modern burrowers, or by examining the functional morphology of their digging organs — the paddle-like forelimbs of moles, or the strengthened snout of burrowing lizards which can be used to batter a tunnel into soil. Another approach, which has now yielded much new information, is to examine fossil burrow systems.

The commonest fossil vertebrate burrows are those ascribed to lungfish. Dozens of examples have now been reported in rocks dating from the Devonian (360–408 million years ago) to the present day. The modern African lungfish, *Protopterus*, burrows into the mud of drying lakes in summer to aestivate — the metabolic rate slows down, and the

animal survives until the rains fall. The aestivation burrow is up to 1 m deep, and the lungfish curls itself up into a ball and surrounds itself with a mucus envelope to prevent the loss of body water. There is only a small opening at its mouth to allow breathing. Only a few fossil lungfish burrows contain the bones of lungfish^{1–3}, but the shape of the burrows and their relationship to the sediment indicate their origin (Fig. 1).

Cylindrical upright pipe-like structures, 3–25 cm in diameter and up to 2 m deep, cutting through several layers of sandstone and mudstone, are typical of burrows from the early Devonian^{4,5}. The burrows taper downwards to a rounded base. It is not certain that these burrows were made by lungfish, but if they were, aestivation behaviour must have arisen very early, as the earliest skeletal fossils date from the same time. Postulated lungfish burrows have often been reported from rocks younger than the Devonian^{1–3,6}. There have been changes in size over time — post-Devonian specimens rarely exceed 15 cm in diameter and 1.7 m in depth (typical sizes are 5–10 cm diameter, 25–50 cm depth). They frequently occur in

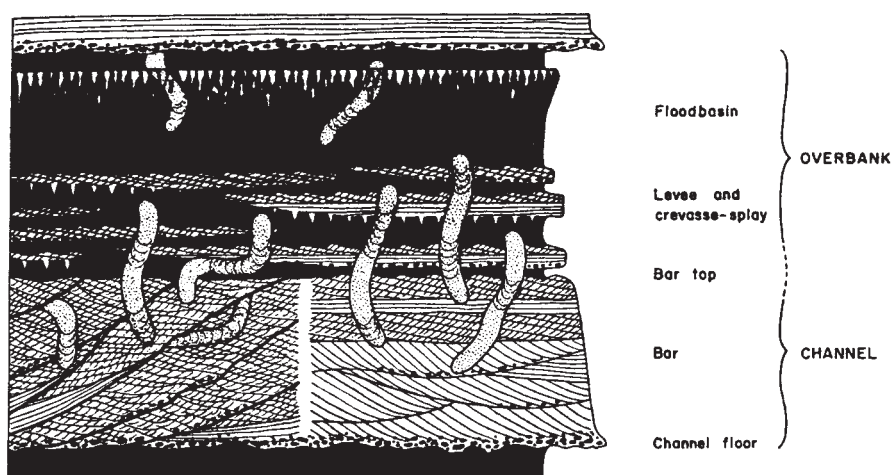


Fig. 1 Postulated lungfish burrows from the early Devonian of South Wales. The surrounding sediments (black, mudstones; cross-hatching, sandstones; white, calcareous soil horizons) indicate warm-flood plain conditions. (From ref. 4.)

large assemblages, and generally in sediments that indicate warm conditions and occasional arid periods.

Smith has recently reported⁷ the surprising discovery of well-preserved burrow systems made by dicynodonts, an extinct vertebrate group (Fig. 2). He found large burrows in the late Permian (240-million-year-old) Beaufort Group of South Africa. The burrows are helical in shape, starting with a modest spiral about 6 cm in diameter at the top, expanding to about 16 cm at the base and then straightening out into a 25-cm-broad section. In all 50 or so specimens, the helix is dextral (similar finds from the Northern Hemisphere are awaited with interest!). The burrows are clearly set off from the surrounding sandstones, being generally filled with mudstone or siltstone that would have entered from above after the burrow went out of use.

The upper shaft of the burrow slopes down at an angle of about 30–35°, and the overall depth of the lower chamber is 0.5–0.75 m from the former land surface. The

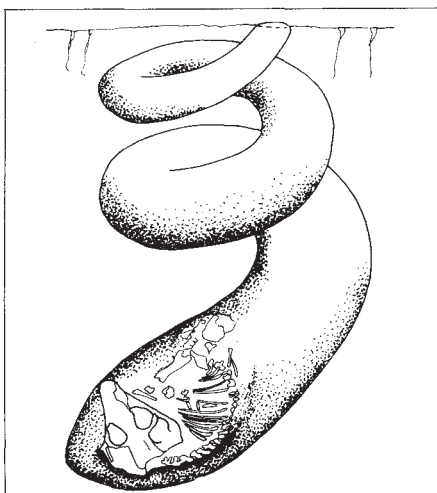


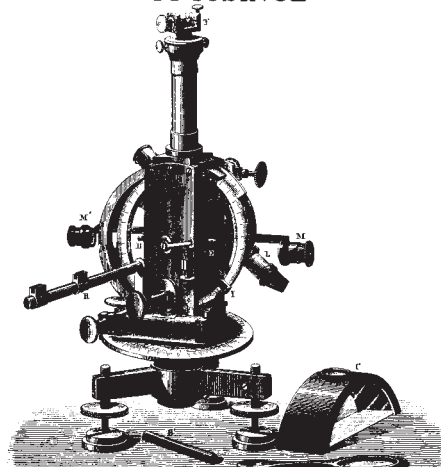
Fig. 2 The dicynodont *Diictodon* preserved curled-up in the terminal chamber of its spiral living burrow, from the late Permian of South Africa. (Based on information in ref. 7.)

walls of the burrows, particularly those of the lower chamber, are marked by irregular scratch marks. The sediments in which the burrows occur are thought to be from an ancient flood plain in which large meandering rivers deposited sandstones, and periodic floods swept finer sediments over the area.

What animals produced these burrows? Large burrows of this kind had hitherto been ascribed only to mammals, which had not evolved by the late Permian. But some remarkable specimens contain skeletons of small mammal-like reptiles, identified as the dicynodont *Diictodon*, which was common at that time. The skeletons are complete, and were found in a curled-up attitude in the terminal chamber of the burrows, which is where *Diictodon* may have sheltered from the heat. The animals excavated the burrows by scratching with the feet and possibly the beak, and the depth was probably limited by the level of the water table.

Similar helical burrows are well known in much younger sediments, generally of Miocene age (5–24 million years ago). *Daimonelix*^{8–10} burrows, on the basis of fossil evidence and comparisons with modern burrows, are thought to have been produced by rodents. Typical *Daimonelix* burrows occur in many parts of the world, ranging in age from the Lower Eocene to the Miocene (55 million years ago), but the best examples come from the late Oligocene to early Miocene Harrison Formation of Nebraska¹¹. Here, the burrows have an upper entrance pit, a middle vertical spiral and a lower living chamber. However, these later forms may have as many as 20 whorls and they can coil dextrally and sinistrally in the same locality. Further, the tube diameter remains constant and does not increase with depth. The Miocene burrows have been confidently ascribed to *Palaeocastor*, a primitive beaver, on the basis of complete and incomplete skeletons found in the living chamber, which is thought to

100 years ago A NEW MAGNETIC SURVEY OF FRANCE



In some respects the magnetometer employed by M. Moureaux possesses advantages over the Kew pattern. In the matter of weight alone there is a considerable difference. A Kew magnetometer weighs nearly 50 pounds, whereas that of the French observers weighs only about 9 pounds. A further advantage possessed by the French model is that it is also an altazimuth instrument, and hence the observer is less dependent upon the knowledge of true time.

From *Nature* 37, 247; 12 January 1888.

have been made for breeding¹¹.

The recent discoveries show that burrowing activities of vertebrates have a long history. Analogy with living forms, together with palaeoenvironmental interpretations, suggest why the burrows were made. Fossils provide evidence for the existence of previously unsuspected, complex burrowing behaviours in completely extinct groups, such as the small dicynodonts, long before the first appearance of rodents in the fossil record. The fossils can also elucidate aspects of behavioural evolution in vertebrates. The Eocene and Oligocene *Daimonelix* burrows, for example, are much simpler than those from the Miocene, and are more like the Permian *Diictodon* burrows, suggesting that the burrow-forming behaviour of rodents has changed through time. The burrow morphology gives hints of how this happened. The fossil lungfish burrows also show that the overall size of this species decreased through time. □

1. Romer, A.S. & Olson, E.C. *Breviora* 30, 1 (1954).
2. Carlson, K.J. *J. Geol.* 76, 641 (1968).
3. Olson, E.C. & Bolles, K. *Fieldiana (Geol.)* 33, 271 (1975).
4. Allen, J.R.L. & Williams, B.T. *Geol. J.* 16, 255 (1981).
5. O'Sullivan, M.J., Cooper, M.A., MacCarthy, I.A.J. & Forbes, W.H. *J. geol. Soc. Lond.* 143, 897 (1986).
6. Dubiel, R.F. *et al. J. sed. Petrol.* 57, 512 (1987).
7. Smith, R.M.H. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 60, 155 (1987).
8. Barbour, E.H. *Science* 19, 99 (1892).
9. Barbour, E.H. *Bull. geol. Soc. Am.* 8, 305 (1897).
10. Schultz, C.B. *Univ. Nebr. Stud. Sci. Technol.* 2, 1 (1942).
11. Martin, L.D. & Bennett, D.K. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 22, 173 (1977).

M. J. Benton is in the Department of Geology, Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland.

Animal behaviour

Dumping eggs on conspecifics

N.B. Davies

THE curious habits of the cuckoo (*Cuculus canorus*) were known to Aristotle^{1,2}, who wrote about 2,300 years ago: "it lays its eggs in the nest of smaller birds . . . they do not sit, nor hatch, nor bring up their young". Gilbert White³ regarded this parasitic habit as a "monstrous outrage on maternal affection", and he dissected a cuckoo to see whether the crop lay unusually low in the body so as to make incubation uncomfortable. Finding no difference in structure from species which incubated their own eggs, he remarked "we are still at a loss for the cause". Since Darwin, naturalists have found the cuckoo's habit less puzzling and have wondered instead why more animals do not trick others into raising their offspring

various populations of starlings (*Sturnus vulgaris*)⁵, up to 24 per cent in colonies of cliff swallows (*Hirundo pyrrhonota*)⁷, 3–31 per cent in swallows (*H. rustica*)¹⁰, and 18 per cent in a population of moorhens (*Gallinula chloropus*)⁸. Presumably, many instances of parasitism have gone undetected, and the use of DNA fingerprinting may reveal these figures as underestimates. In various species of ducks, the percentage may exceed 50 per cent of nests parasitized⁵: in one study¹¹ it reached 95 per cent.

Some of these high frequencies of parasitism have been recorded in populations breeding at unusually high densities because of the provision of nest boxes¹¹. The nests parasitized are also often those

In cliff swallows, as in other species^{5,8}, parasitic females adopt a mixed strategy of raising some of their own young and dumping some eggs parasitically. It will be interesting to see whether different females show varying degrees of parasitic behaviour and whether any, like the cuckoo, are professional parasites. The effects of brood parasitism on both host and parasite reproductive success need to be quantified. In some species, hosts can recognize conspecific eggs which are not their own and eject them from the nest^{14,15}, whereas in others eggs are ejected only if they appear in the nest before the host begins to lay — after this, parasitic eggs are accepted^{10,16,17}.

Brown and Brown suggest⁶ that dumping eggs is the way a bird avoids putting all its eggs in one basket; in a risky environment spreading eggs in several nests may increase the chance that at least some young survive. However, recent theoretical models suggest¹⁸ that selection for this form of bet-hedging is weak. A more

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Changelings in the nest — the monstrous young cuckoo being fed by a dutiful yellow wagtail (left), is easily identified, but it is impossible to see whether one of a brood of cliff swallows (right) is an intruder. (Cuckoo: Frank Lane Picture Agency. Cliff swallows courtesy of Charles and Mary Brown.)

and so avoid the costs of parental care. Recent observations^{4,5} show that brood parasitism is, in fact, much commoner than previously supposed. Within several species of birds, some individuals play at cuckoos and lay eggs in the nests of conspecifics. On page 66 of this issue⁶, Charles and Mary Brown show, for the first time, that intraspecific brood parasitism can take place not only by direct laying of eggs in other nests, but also by the transfer of partly incubated eggs from the parasite's own nest.

Brood parasitism within a species has been detected by direct observations⁷ of females laying in other individuals' nests and also, indirectly, by the appearance of more than one egg per day during the laying of a clutch¹⁷ (each female lays only one egg per day); by the occurrence of eggs different in shape and colour from those of the host female⁸ (in some species each female lays distinctive eggs); and by maternity exclusion using electrophoresis⁹. The proportion of nests parasitized has been estimated as 5–46 per cent for

of near neighbours^{7,10}. This result raises the question of whether laying eggs in other nests arises simply as mistaken nest identity rather than as an adaptation favoured by natural selection. The observation reported in this issue⁶ by Brown and Brown of female cliff swallows transporting partly incubated eggs from their own nests to the nests of other females supports the second interpretation, at least for this species; such specialized behaviour seems hardly likely to represent a mistake. Furthermore, intraspecific parasites adopt other tactics which seem to increase the success of parasitism. Both starlings¹² and cliff swallows⁶ sometimes remove host eggs before depositing their own, which may result in their offspring receiving better parental care. Parasitic laying in starlings occurs later in the day than the peak period of laying¹³, when the hosts are more likely to be at their nests and able to evict intruders. Finally, parasites can lay their eggs remarkably quickly⁷. Interestingly, all these last three tactics are also adopted by cuckoos.

likely explanation, perhaps, is that parasites increase their lifetime reproductive success by reducing the costs of parental care. □

1. Aristotle (ed. Peck, A.L.) *Historia Animalium* Vol. II (Heinemann, London, 1970).
2. Aristotle (ed. Hett, W.S.) *On Marvellous Things Heard* (Heinemann, London, 1936).
3. White, G. *The Natural History of Selborne* (White, London, 1788–89).
4. Yom-Tov, Y. *Bio. Rev.* **55**, 93–108 (1980).
5. Andersson, M. in *Producers and Scroungers: Strategies of Exploitation and Parasitism* (ed. Barnard, C.J.) 195–228 (Croom Helm, London, 1984).
6. Brown, C.R. & Brown, M.B. *Nature* **331**, 66–68 (1988).
7. Brown, C.R. *Science* **224**, 518–519 (1984).
8. Gibbons, D.W. *Behav. Ecol. Sociobiol.* **19**, 221–232 (1986).
9. Gowaty, P.A. & Karlin, A.A. *Behav. Ecol. Sociobiol.* **15**, 91–95 (1984).
10. Möller, A.P. *Anim. Behav.* **35**, 247–254 (1987).
11. Semel, B. & Sherman, P.W. *Auk* **103**, 813–816 (1986).
12. Feare, C.J. *The Starling* (Oxford University Press, 1984).
13. Feare, C.J., Spencer, P.L. & Constantine, D.A.T. *Ibis* **124**, 174–178 (1982).
14. Bertram, B.C.R. *Nature* **279**, 233–234 (1979).
15. Victoria, J.K. *Ibis* **114**, 367–376 (1972).
16. Emlen, S.T. & Wrege, P.H. *Ethology* **71**, 2–29 (1986).
17. Stouffer, P.C., Kennedy, E.D. & Power, H.W. *Anim. Behav.* **35**, 1583–1584 (1987).
18. Bulmer, M.G. *J. theor. Biol.* **106**, 529–535 (1984).

N.B. Davies is in the Zoology Department, University of Cambridge, Cambridge CB2 3EJ, UK.

Cell biology

Calpactin in exocytosis?

Robert D. Burgoyne

THE exocytotic fusion of membrane vesicles with the plasma membrane probably occurs in all cells and is involved, for example, in the insertion of new components into the plasma membrane and in the release of secretory products from vesicles. In the case of cells specialized to secrete in a regulated fashion, exocytosis is triggered on cell activation by an intracellular signal which is often a rise in the concentration of cytosolic calcium. The importance of calcium in the regulation of exocytosis has led to a search for putative calcium receptors involved in mediating the interaction and fusion of secretory vesicles with the plasma membrane. A report by Drust and Creutz on page 88 of this issue¹ points to a possible role for the protein calpactin in the calcium-dependent control of membrane fusion.

In attempts to identify calcium-regulated proteins that could be involved in exocytosis, several proteins were identified in adrenal medullary extracts that could be isolated because of their ability to bind reversibly to secretory vesicle membranes in the presence of micromolar concentrations of calcium^{2,3}. Among these proteins are calmodulin, protein kinase C (ref. 4), caldesmon⁵ and a series of closely related calcium- and lipid-binding proteins, now known to be widely distributed, which have been given the generic name annexins⁶. Earlier data from protein and complementary DNA sequence analysis had shown significant similarity between some of the annexins⁷, and further recent results⁸⁻¹⁰ show that there are at least five separate members of this family (calpactin, lipocortin I, endonexin, endonexin II and p70) which share conserved domains. Calpactin differs from other annexins in being a heterotetramer with the structure p36₂ p10₂. The heavy chain of calpactin (p36) was already known as a major substrate for phosphorylation by retroviral tyrosine kinases.

One test of the possible significance of calcium-dependent binding of annexins to membranes is whether they bind at physiologically relevant calcium levels. The levels of calcium required for binding of annexins to membranes or liposomes are variable⁷ (see table) but it is likely that all these proteins become associated to some extent with membranes following cell activation and elevation of calcium to low micromolar levels. Drust and Creutz¹ address the additional question of whether any of the annexins mediate membrane-membrane interactions in stimulated cells. Their data (see ref. 13 for data on p70) show that only calpactin is able

to aggregate secretory granules at low micromolar levels of calcium, and that the calcium-dependency is similar to that of calcium-activated secretion in permeabilized chromaffin cells. Truncated forms of the calpactin heavy chain, which are unable to bind p10, show much higher calcium requirements for granule aggregation¹. Powell and Glenney¹¹ have also recently found that intact calpactin will aggregate liposomes at lower calcium levels than the isolated heavy chain. These results suggest that the ability of calpactin to aggregate secretory granules or liposomes at much lower calcium levels than the other annexins may be conferred by

use in finally determining the importance of calpactin in the permeabilized cell. Chromaffin cells permeabilized by treatment with digitonin will secrete in response to addition of micromolar levels of calcium. If after permeabilization, however, the cells are maintained at low calcium levels, they eventually leak proteins essential for exocytosis, and secretory responsiveness can be reconstituted by adding back the leaked proteins¹⁵. The factor(s) lost from the cells cannot be replaced by re-addition of calmodulin, nor can loss of secretory responsiveness be prevented by treating the cells with phorbol ester to reduce leakage of protein kinase C. The conditions used for incubation of the cells at low calcium are such that it is likely that calpactin and other annexins would be among the leaked proteins. An examination of the effect of calpactin on cells that have become refractory to calcium stimulation because

Calcium-dependence of annexin binding to membranes or liposomes

Protein	Concentration of Ca ²⁺ for half-maximal binding (μM)	Ref.
p70*	2.6	2
lipocortin [†]	4 [§] , 20	17
endonexin I (32K calelectrin)*	5.5	16
endonexin II [†]	40	9
calpactin [†]	<0.01 [‡] , 5	11

The table lists the five distinct proteins that have been shown to be members of the annexin family on the basis of their possession of conserved common domains. For alternative names that have been given to these proteins, see ref. 6.

*Binding assayed with chromaffin granule membranes.

[†]Binding assayed with phosphatidylserine liposomes.

[‡]Binding assayed with phosphatidylethanolamine liposomes.

[§]Data for phosphorylated form.

[‡]A large component of binding was detected at low calcium levels, with a second calcium-dependent component.

p10. The association of p10 could increase the affinity of the calcium and/or lipid-binding sites within the heavy chain or, alternatively, the heterotetrameric structure of intact calpactin could facilitate membrane crosslinking.

Is calpactin the calcium receptor in exocytosis? The results of Drust and Creutz show that at least in an *in vitro* model of membrane interaction calpactin can mediate membrane-membrane association, which in the presence of Mg²⁺ and arachidonic acid results in membrane fusion. There is as yet no direct evidence that such a mechanism occurs during exocytosis. Calpactin is localized in the subplasmalemmal region of chromaffin cells¹² and shows increased incorporation of radiolabelled phosphate on stimulation¹³, which could indicate a phosphorylation-dependent mechanism for inactivation of calpactin¹¹ following exocytosis. Also arachidonic acid is produced on stimulation in chromaffin cells¹⁴, so it is possible that exocytotic fusion in these cells involves the same components that are effective in the *in vitro* experiments.

An experimental model likely to be of

of protein leakage may well help to determine whether this protein is indeed involved in exocytosis. □

1. Drust, D.S. & Creutz, C.E. *Nature* **331**, 88-91 (1988).
2. Geisow, M.J. & Burgoyne, R.D. *J. Neurochem.* **38**, 1735-1741 (1982).
3. Creutz, C.E. *et al. J. Biol. Chem.* **258**, 14664-14674 (1983).
4. Summers, T.A. & Creutz, C.E. *J. Biol. Chem.* **260**, 2437-2443 (1985).
5. Burgoyne, R.D., Cheek, T.R. & Norman, K.M. *Nature* **319**, 68-70 (1986).
6. Geisow, M.J., Walker, J.H., Boustead, C. & Taylor, W. *Biosci. Rep.* **7**, 289-298 (1987).
7. Geisow, M.J. & Burgoyne, R.D. *Ann. N.Y. Acad. Sci.* **493**, 563-576 (1987).
8. Weber, K. *et al. EMBO J.* **6**, 1599-1604 (1987).
9. Schlaepfer, D.D., Mehlman, T., Burgess, W.H. & Haigler, H.T. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6078-6082 (1987).
10. Martin, F., Derancourt, J., Capony, J.-P., Colote, S. & Cavadore, J.-C. *Biochem. biophys. Res. Commun.* **145**, 961-968 (1987).
11. Powell, M.A. & Glenney, J.R. *Biochem. J.* **247**, 321 (1987).
12. Burgoyne, R.D. & Cheek, T.R. *Biosci. Rep.* **7**, 281 (1987).
13. Creutz, C.E. *et al. J. Biol. Chem.* **262**, 1860-1868 (1987).
14. Frye, R.A. & Holz, R.W. *J. Neurochem.* **43**, 146-150 (1984).
15. Sarafian, T., Aunis, D. & Bader, M.-F. *J. Biol. Chem.* (in the press).
16. Sudhof, T.C. *et al. Biochemistry* **23**, 1103-1109 (1984).
17. Schlaepfer, D.D. & Haigler, H.T. *J. Biol. Chem.* **262**, 6931-6937 (1987).

Robert D. Burgoyne is in the MRC Secretory Control Research Group, The Physiological Laboratory, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK.

Competitive elimination of neuromuscular synapses

SIR—Callaway *et al.*¹ recently concluded that inactive motor neurons have an advantage over active motor neurons during neuromuscular synapse elimination in neonatal rabbit skeletal muscle. It is important to draw attention to certain limitations in their methodology and to exercise some caution in the interpretation of the data.

First, Callaway *et al.* report data obtained using only a single, indirect measure of innervation: twitch contractions of isolated skeletal muscles in response to stimulation of individual motor axons. It is well known that twitch tension measurements can overestimate both the amount of polyneuronal innervation and motor unit size, even though extreme precautions may be taken to ensure that the recordings are isometric². Tetanic tension measurements are more reliable because time is given during a tetanus for series compliance in muscle to be saturated. Corroborative measurements, using intracellular recording or measurement of motor unit size using selective glycogen depletion of motor units, and histology³ are also necessary to support the contention.

Second, Callaway *et al.* base their conclusions on differences in the sizes of the twitch tensions produced only by the inactive motor units. There are two potential problems with interpretation here that could be addressed using direct methods. One is that, under some conditions, chronically paralysed neuromuscular junctions respond repetitively to single shock stimuli to the nerve, leading to summation of twitch tension, and therefore to spurious overestimates of motor unit size. This has been demonstrated using intracellular recording⁴. A second problem is that synapse elimination is delayed in muscle fibres which become innervated only by convergent terminals from separate tetrodotoxin-blocked motor neurons⁵. At any particular stage in the developmental process this would produce larger than normal inactive motor units, but no difference in the size of active units in the same muscles, as Callaway *et al.* actually report.

Finally, Callaway *et al.* make an important, additional observation which should also be confirmed using direct methods: the measurements of tension overlap indicate that some muscle fibres become innervated exclusively by inactive motor axons. This does not mean that inactive axons must have had a competitive advantage. That conclusion could only be made if there had been a corresponding decrease in the mean size of the active motor units in the muscles. A simpler explanation is that factors other than

differences in activity contribute to the selective stabilization of terminals; indeed it would be surprising if it were not so.

That muscle paralysis or stimulation influences the overall rate of synapse elimination is generally accepted⁵⁻⁸. In relation to synaptic competition however, a cautious conclusion, based on the study of Callaway *et al.* and a complementary one by Ridge and Betz⁹, is that it is difficult to demonstrate a clear effect of differences in motor neuron activity during synapse elimination in immature skeletal muscle. Furthermore, such differences, if they occur, are small and of uncertain functional significance for the maturation of motor innervation patterns. Even during nerve regeneration, where the effects of inactivity during competition seem clear cut^{4,10}, differences in motor neuron activity cannot be an overriding influence in the interactions between motor neurons during synapse formation and elimination.

R. R. RIBCHESTER

Department of Physiology,
University Medical School, Teviot Place,
Edinburgh EH8 9AG, UK

1. Callaway, E.M., Soha, J.M. & Van Essen, D.C. *Nature* **328**, 422-426 (1987).
2. Brown, M.C. & Matthews, P.B.C. *J. Physiol., Lond.* **151**, 436-457 (1960).
3. Jones, S.P., Ridge, R.M.A.P. & Rowlerson, A. *J. Physiol., Lond.* **386**, 377-384 (1987).
4. Ribchester, R.R. & Tait, T.J. *J. Physiol., Lond.* **344**, 89-111 (1983).
5. Thompson, W.J., Kuffler, D.P. & Jansen, J.K.S. *Neuroscience* **4**, 271-281 (1978).
6. Thompson, W. *Nature* **302**, 614-616 (1983).
7. O'Brien, R.A.D., Östberg, A. & Vrbová, G. *J. Physiol., Lond.* **282**, 571-582 (1978).
8. Tait, T.J. *J. Physiol., Lond.* **340**, 175-194.
9. Ridge, R.M.A.P. & Betz, W.J. *J. Neurosci.* **4**, 2614-2620 (1984).
10. Ribchester, R.R. in *Electrical stimulation and neuromuscular disorders* (eds Nix, W. & Vrbová, G.) 53-63 (Springer, Heidelberg, 1986).

CALLAWAY *ET AL.* REPLY — Ribchester raises four specific concerns about our experimental paradigm and the interpretation of our results. First, adequacy of twitch tension measurements. We did not use tetanic tension measurements, because preliminary experiments showed that this led to gradual irreversible tension fatigue in our *in vitro* preparation. The key issue is not whether twitch tensions are less accurate than tetanic tensions as a measure of absolute motor unit size but whether they provide a reliable measure of changes in motor unit size in normal and experimental muscles. We agree that lack of perfect isometric recording conditions means twitch tensions tend to overestimate absolute motor unit size¹, but the recording conditions are similar for all motor units from a given muscle. Twitch tensions therefore provide a reasonable indication of the relative size of different units in the same muscle.

We emphasize that our primary assay compared the relative tensions of motor unit populations from different spinal

roots in the same animal. This within-animal normalization procedure minimized the sensitivity of our analysis to any systematic tendency of twitch tensions to overestimate the percentage of muscle fibres in a motor unit. We would welcome any corroborative measurements using other procedures but we contend that the high statistical significance of our major findings demonstrates the sensitivity and reliability of our paradigm.

Second, repetitive nerve firing. Our estimates of twitch tension for inactive motor units would have been systematically overestimated if there was repetitive nerve firing in this subpopulation of motor axons. Inactivity-induced repetitive firing has been reported² but only where much of the muscle was inactive; in these cases motor-unit twitch rise times were significantly slowed. Our paradigm was designed to maintain activity over nearly all the muscle and thus should not have been subject to this effect. In support, we found that the average rise time of motor units from tetrodotoxin (TTX)-inactivated roots was within 2 per cent of that for motor units from active roots in the same animals and for those of control and normal animals (fast and slow motor-unit rise times analysed separately).

Third, delayed synapse elimination. We presented evidence against a delay in synapse elimination at end-plates jointly innervated by active and inactive inputs. Ribchester suggests instead that the delay might have been restricted to end-plates that had lost their final active input but still received multiple inputs from inactive motor neurons. Such fibres should have been rare in our experiments and are unlikely to have contributed significantly to the larger size of inactive motor units. We have recently obtained more direct evidence against this possibility by allowing 5-7 days of recovery after 4-5 days of TTX inactivation. The previously inactivated motor units remain significantly larger than their active counterparts in control animals, by an amount comparable to that found in the short-term experiments (in preparation). Moreover, analysis of silver-stained end plates confirms that there is no significant degree of polyinnervation in these muscles. As well as arguing strongly against the delay hypothesis, these observations provide further evidence against the concern about repetitive firing, because the preparation had 5-7 days of presumably normal activity to recover from any side effects of inactivity.

Finally, decrease in size of active motor units. If inactive motor units have an advantage, the active motor units against which they compete should indeed have become smaller than normal. The magnitude of the expected difference is

however, rather small because the inactive motor units were very much in the minority. Their substantially larger sizes would be balanced by a smaller change in size among a larger population of active motor neurons. Moreover, detection of any such difference entails a comparison of motor-unit tensions scaled to the maximum direct tension — without using the within-animal normalization we employed to circumvent the significant individual variability in average motor unit sizes found even in normal animals. Nonetheless, a small bias in the appropriate direction was reported in our initial sample. Including the additional data obtained in the long-term recovery experiments (see above) makes this difference statistically significant ($P < 0.05$).

In summary, we feel we have now demonstrated a clear effect of differential activity on neuromuscular competition, and we hope that we have adequately addressed any residual doubts. We concur with Ribchester that differential activity is not the only factor contributing to the selective stabilization of synapses but it is likely to be important, given that the effects we observed were quite substantial (> 45 per cent), and particularly because of the short time for which differential activity was maintained. Effects of this magnitude, if integrated over the full period of synapse elimination, could in principle make an important contribution to sculpting motor-unit size distributions during normal maturation.

EDWARD M. CALLAWAY
JAMES M. SOHA
DAVID C. VAN ESSEN

Division of Biology 216-76,
California Institute of Technology,
Pasadena, California 91125, USA

1. Brown, M.C. & Matthes, P.B.C. *J. Physiol., Lond.* **151**, 436–457 (1960).
2. Ribchester, R.R. & Tait, T.J. *J. Physiol., Lond.* **334**, 89–111 (1983).

Sequence patterns in protein kinases

SIR—Following Brenner's report¹ on phosphotransferase homologies, we have designed two sequence patterns that can be used to identify protein kinases and to discriminate them from other proteins. The first of these, derived from the one proposed by Brenner, is (LIV)(HY)XD (FILMVY)XXXNX(FILMV)(FILMV).

The second, much shorter and less degenerate, is (LIV)GXGX(FY)GX(LIV) and is based on alignments of protein kinase sequences in the vicinity of a lysine residue that has been shown² to be involved in ATP binding in both oncogenic tyrosine kinases and cyclic (c)AMP-dependent protein kinases.

(For both patterns the amino acids in parentheses are alternatives; X stands for any amino acid.)

FPHDWDLFKLMGIGSGGFGPVFESGFIEIL
:::.. ::::: ::::: ::::: ::::: :::::
FPREKLTIRLL-LGSGAFGEVYEGTAVDIL

TMS1 (424–453)

Human *ros* (14–42)

Sequence comparison of TMS1 and human *ros*. Identical amino acids are indicated by :, different but conserved amino acids by ., and the sequence pattern is underlined.

It must be noted that Brenner's pattern, while identifying most known phosphotransferase sequences, only flags a subset of all known protein kinases. For instance, it fails for cAMP- and cGMP-dependent kinases, protein kinases C and the *abl*, *fes*, *fps*, *ros* and *trk* oncogene products.

The second pattern is a generalization of the LGXGXFGXV pattern first pointed out by Barker and Dayhoff as a conserved element between the *src* gene product and the bovine cAMP-dependent protein kinase³.

These patterns have been designed and their specificity tested by scanning the SWISS-PROT 5.1 databank (5,300 protein sequences, 1,400,000 residues) using the PC/Gene sequence analysis package (developed by A. Bairoch and D. Gavin, University of Geneva, and distributed by Genofit SA and IntelliGenetics Inc.), as well as the PseqIP 3.0 databank⁴ (6,487 protein sequences, 1,700,000 residues) using the SASIP sequence analysis package⁵. A 'reverse translation' of the second pattern was also used to scan GenBank 50.0 (ref. 6) and the EMBL 10.0 nucleotide sequence data library⁷.

Table 1 Sequence patterns used to identify protein kinases and discriminate them from other proteins

Protein family	Pattern 1	Pattern 2	Kinase type
cAMP-dependent PK	+	+	ST*
cGMP-dependent PK	+	+	ST
Phosphorylase <i>b</i> kinase	+	–	ST
Protein kinase C	+	+	ST
CDC2/CDC28	+	+	?
CDC7	+	–	?
PK-25/YKR	+	+	?
HSV kinase related	+	–	?
Myosin light chain kinase	–	+	?
EGF receptor (<i>erb-B</i> , <i>neu</i>)	+	+	Y†
Insulin receptor	+	+	Y
<i>abl/dash</i>	+	+	Y
<i>fes/fps</i>	+	+	Y
<i>fgr</i>	+	+	Y
<i>fms</i>	+	+	Y
<i>kit</i>	+	+	Y
LSK-T/ <i>tck</i>	+	+	Y
<i>met</i>	+	+	Y?
<i>mill/mhl/ras</i>	+	+	ST
<i>mos</i>	+	+	ST
<i>pim-1</i>	+	+	?
<i>ros</i>	+	+	Y
<i>src/yes/syn</i>	+	+	Y
<i>trk</i>	+	+	Y

* Serine/threonine; † tyrosine. References to all sequences are available from the authors.

As shown in the accompanying table, the first pattern found every known protein kinase except the myosin light-chain kinases, while the second, shorter pattern flagged all but the gamma chain of phosphorylase *b* kinase, the yeast CDC7 gene product and a kinase-related protein from herpes simplex virus type 1.

Whereas the first pattern did not detect any sequence other than known protein kinases, the second pattern pointed out two other sequences: the hexon-associated protein IX from adenovirus types 7 and 3 (but not the homologous protein from adenovirus types 5 and 12) and the TMS1 protein coded on the Ti plasmid from *Agrobacterium tumefaciens*.

Because of its involvement in plant tumorigenesis, the latter case was investigated further. Using the Fastp program⁸ a good but very localized similarity was detected between the TMS1 protein and the *ros* oncogene product (as well as the mouse *pim-1* oncogene) around the second pattern, as shown in the figure.

We tested other patterns from different conserved regions of protein kinases, but none was as specific as those reported here, which thus seem to be bona fide sequence signatures for protein kinases. Accordingly, it is tempting to propose that the *A. tumefaciens* TMS1 tumorigenic gene product exhibits protein kinase activity.

AMOS BAIROCH

Department of Medical Biochemistry,
University of Geneva Medical School
1211 Geneva 4, Switzerland

JEAN-MICHEL CLAVERIE

Computer Science Unit, Institut Pasteur,
75724-Paris Cedex 15, France

1. Brenner, S. *Nature* **329**, 29 (1987).
2. Kamps, M.P., Taylor, S.S. & Sefton, B.M. *Nature* **310**, 589–592 (1984).
3. Barker, W.C. & Dayhoff, M.O. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2836–2839 (1982).
4. Claverie, J.-M. & Bricault, L. *Proteins* **1**, 60–65 (1986).
5. Claverie, J.-M. & Sauvaget, I. *Cabios* **1**, 95–104 (1985).
6. Bilosky, H.S. *et al. Nucleic Acids Res.* **14**, 1–4 (1986).
7. Hamm, G.H. & Cameron, G.N. *Nucleic Acids Res.* **14**, 5–9 (1986).
8. Lipman, D.J. & Pearson, W.R. *Science* **227**, 1435–1441 (1985).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references. □

The exclusion principle

Valentin F. Turchin

Science, Philosophy, and Human Behavior in the Soviet Union. By Loren R. Graham. Columbia University Press: 1987. Pp.565. \$45, £26.50.

PROFESSOR Graham's book is essentially an updated and expanded edition of his *Science and Philosophy in the Soviet Union*, published in 1972. It starts with a short historical overview (Chapter 1) and a brief discussion of the official Soviet philosophy, dialectical materialism (Chapter 2). The rest of the book (Chapters 3 to 12) contains an account of publications and controversies in the Soviet Union on the borders of philosophy and various fields of science: biology, psychology, cybernetics and computers, chemistry, physics, and cosmology. The infamous deeds of Lysenko and Co. are excluded, because, as the author explains, his main goal was not

the recounting of this repressive side of the story of Soviet science. . . . Instead, I have sought to emphasize the more philosophically interesting, rather than the politically dramatic, aspects of the relation between Soviet philosophy and science [p.433].

The specialized chapters are independent, and the author suggests that the reader interested in a specific field of science jumps, after reading the first two chapters, to the chapter of interest. I base my judgement of the book on the chapters on cybernetics and computers, quantum mechanics, and relativistic physics, subjects that are inside my area of expertise.

The factual material in the specialized chapters is drawn from those writings that were most noted at the time of publication, and that produced most discussion. Graham's book will no doubt be a valuable bibliographical source for a student of the history of science and philosophy in the Soviet Union (bibliographical notes take more than one hundred pages). But the way the author presents the official Soviet philosophy and its interaction with science surprised me beyond any measure. It simply has nothing to do with the reality of things; indeed, it is contradicted by the book's own contents.

According to Soviet textbooks, dialectical materialism is a redefinition of Hegel's dialectical method in terms of philosophical materialism. One can distinguish the following elements in this philosophy: (i) naive realism, which serves as the foundation of the whole construction, and makes it attractive for mass consumption; (ii) traditional metaphysics manipulating a few abstract notions, such as matter, space, time, as distinct entities that logically precede human experience and manifest themselves in it; (iii) a few dialectical schemes,

"negation of the negation" and so on, picked up by Engels from the Hegel heritage; (iv) the idea of a universal evolution.

None of these elements is, of course, original; they were taken from the philosophical thought of the second half of the nineteenth century, when dialectical materialism was conceived. And all of them have some merit — even a dose of naive realism may sometimes be useful. As for evolution, it is one of the greatest ideas in human history. The question which presents itself immediately, but is not even raised by Graham, is why should these elements, chosen almost at random, be announced as the worldview of the Communist party? Why should a great many other ideas of European philosophy, in fact all those that do not fit the Procrustean bed of dialectical materialism, be banned, and their proponents denounced, jailed, eliminated?

The core of the Soviet philosophy which is important for the Communist party is the mixture of naive realism and metaphysics that states that, firstly, there are objective laws of nature, that, secondly, they are cognizable, and, thirdly, that it was 'scientifically proven' by Marx that they necessarily lead to the triumph of Communist revolution all over the world. If ancient monarchs derived their rule from God's will, the Communists derive their legitimacy from 'objective necessity'. Anything that may undermine, in the eyes of the masses, the firm belief in

this trinity, is perceived as subversion. Dialectical materialism is, essentially, a negative concept. Its originality and significance is not in the ideas it insists on; it is in the ideas it excludes and denounces as 'bourgeois' and 'idealistic'.

One of the main sources of dialectical materialism is Lenin's book *Materialism and Empirio-Criticism*. Professor Graham fails to tell the reader what every Soviet textbook on the history of the Soviet Communist Party stresses — the political origins and goals of the book. Lenin rightly saw a threat in Mach's *Analysis of Sensations* and in the whole trend of thought initiated by it, a trend which was supported by some of Lenin's comrades in the party (most notably, Bogdanov). Ever since then the term 'Machist' has been a deadly insult in the Soviet Union, comparable only to 'Trotskyist'. Mach's ideas, however, laid the groundwork for the modern philosophy of science. Soviet official philosophy still lives in the last century. The most important aspect of contemporary philosophy, analysis and definition of existential concepts in terms of human experience, is still a taboo for it. And looming large, as in any dogmatic system, is interpretation of, and comments on, quotations from 'the classics'. As a result, a vast literature comes into existence which poses and 'solves' pseudo-problems.

This is not to say that one can find nothing of interest in Soviet writings on philosophy. The creative spirit of individuals manifests itself under any system of thought. But dialectical materialism cannot take credit for it; for creatively thinking authors it is just a restriction. To have your ideas published, you must put them in such a way that they do not contradict dialectical materialism, at least on the surface. In this situation, it will typically be the authors least recognized

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Directing trends of thought — Lenin during the October Revolution.

officially who have the most interesting things to say. In this context one thinks of such writers as N.F. Ovchinnikov, V.A. Lefebvre and A.A. Zinoviev. But none of these names is in the index of Professor Graham's book. Rather, the author has picked precisely those authors who are most recognized. This is a correct strategy for an historian — but did not Graham vow to be guided instead by philosophical interests?

Given the political origin and restrictive nature of the Soviet philosophical doctrine, it would be a miracle if, in addition to its main function, this doctrine could inspire and guide scientific thought. Graham seems to believe in this miracle. We read in the preface:

I conclude that even good Soviet science bears the mark of Soviet philosophy, including 'hard' sciences, such as physics, a conclusion extremely difficult for most Western scientists to accept [p.xi].

Very difficult, indeed! And not only for Western scientists. The only marks of dialectical materialism that I saw in the Soviet Union were the scars left by cutting away the ideas and theories that were the most characteristic and important of our time. To convince the reader of a miracle, an author must provide really good proof of it. Nothing of the kind can be found in the book. The intellectual landscape of Soviet philosophy, as it appears there, is dreary and barren.

In the specialized chapters, Professor Graham's book consists mainly of summaries of various books and journal papers. I will pick out a few passages to illustrate the underlying ideas and their coverage, choosing those Soviet authors whom Professor Graham himself finds the more significant.

In the chapter on cybernetics we read:

I.B. Novik was one of the more energetic Soviet philosophers who attempted to define information in terms of dialectical materialism. In his book *Cybernetics: Philosophical and Sociological Problems* Novik tried to present a systematic treatment of cybernetics from the standpoint of enlightened Marxism. From the

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Dialectics in progress — Engels and Marx in conversation.

outset he aligned himself with the partisans of cybernetics; he insisted that there was no conflict between this new field and dialectical materialism. Novik explained [my emphasis, VT] that cybernetic information is a property of matter, a property directly connected with Lenin's copy theory of epistemology [p.283].

Hardly an illuminating explanation.

The book by D.A. Ursul, *The Nature of Information*, is characterized as "interesting", and almost two pages are given to a summary of it. But I was unable to catch any new idea deserving of the name.

... Ursul ... defended very strongly his belief that information is characteristic of all matter, from the simplest inorganic forms to human society. Ursul tied this conception of the unity of nature closely to dialectical materialism, arguing that the dialectical laws help one to understand information processes. But he also thought that information theory had added new content to dialectical materialism; he posed the possibility of making a few changes in Marxist philosophy as a result of contribution to man's knowledge of information theory. In particular, he believed that there was a good argument for converting the concept of information from a scientific-technical one to a general philosophic category, adding it to the existing list of categories in the Marxist dialectic [p.284].

A promotion, so to say. Quite a problem for a philosopher.

The chapters on physics recount the history of the controversies over quantum mechanics and relativity theory, in which, roughly, the physicists stood for the standard interpretation of these theories, while the philosophers were against it, though there were defectors on both sides. In fact, both theories, and especially quantum mechanics in the Copenhagen interpretation (accepted now by practically every physicist in the world), are incompatible with the spirit of dialectical materialism, and official Soviet philosophers knew this very well. The physicists, therefore, were looking for a compromise that could allow them to discuss and teach intelligibly the basic concepts of modern physics. Vital though the problem was for the Soviet scientists, it remained a pseudo-problem for the

world at large, and the discussions in the Soviet Union were of little, if no, importance for the philosophy of science.

Of the philosophers, the most attention is given in the chapters on physics to M.E. Omel'ianovskii, who is characterized as "one of the most influential Soviet philosophers of science in the sixties and seventies" (pp.343–344). He is an interesting figure in that he tried to work out the compromise the scientists wanted. Nevertheless — or, rather, because of this — one will fail (as I remember I did back in Russia) to find a single fresh idea in his writings. So I tried to find in Graham's exposition what I may have missed some time ago. I failed again. Here is a typical excerpt:

Determinism is perfectly compatible, according to Omel'ianovskii, with statistical laws. Furthermore, Omel'ianovskii considered the statistical laws of quantum mechanics to be not the result of the uncontrollable influence of measurement (Heisenberg), nor the result of indeterminism governing the individual micro-object (Reichenbach), nor the result of hidden parameters (Bohm), nor the result of the relationship of the microensemble and its macroenvironment (Blokhintsev), but instead the result of what he called the "peculiar wave-corpuscular properties of micro-objects" [p.346].

So, arsenic is poisonous because it has "peculiar" poisonous properties. If not for Molière, it would be a brave new idea.

The concluding remarks in Graham's book start as follows:

Contemporary Soviet dialectical materialism is an impressive intellectual achievement. The elaboration and refinement of the early suggestions of Engels, Plekhanov, and Lenin into a systematic interpretation of nature is the most original intellectual creation of Soviet Marxism. In the hands of its most able advocates, there is no question but that dialectical materialism is a sincere and legitimate attempt to understand and explain nature. In terms of universality and degree of development, the dialectical materialist explanation of nature has no competitors among modern systems of thought [p.429].

Nothing in the book, nor in the actual state of the Soviet philosophy, justifies this impassioned panegyric. Even though the author is a Westerner, the situation is typically and sadly Soviet: you roam around the shops with half-empty shelves; then you open a newspaper and read about the great achievements of the socialist economy. Fortunately, with Gorbachev's *perestroika* the newspapers are becoming more modest. This is a good time for the foreign friends of dialectical materialism to have a small *perestroika* of their own. □

Valentin F. Turchin, a physicist and computer scientist, emigrated from the Soviet Union in 1977. He is now Professor of Computer Science at the City College, City University of New York, Convent Avenue and 138th Street, New York, New York 10031, USA, and is author of *The Phenomenon of Science (1977)* and *The Inertia of Fear and the Scientific Worldview (1981)*.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Hegel — a dialectical heritage.

Avian selection

Tim Halliday

Naturalized Birds of the World. By Sir Christopher Lever. Illustrations by Robert Gillmor. *Longman, Harlow, UK/Wiley, New York: 1987. Pp.615. £65, \$197.*

Extinct Birds. By Errol Fuller. *Viking-Rainbird, London/Facts on File, New York: 1987. Pp.256. £20, \$35.*

THE impact of human civilization on the world's avifauna has been mixed. The great majority of species have declined in numbers, primarily as the result of habitat destruction, and there has been a dramatic increase over the past 300 years in the rate at which birds have become extinct. A few species, however, have prospered as a result of human activities and have been able to establish breeding populations in parts of the world where they did not previously exist.

Such 'naturalized' species are the subject of Christopher Lever's authoritative book. Lever's aim was to document, for every naturalized species in the world, exactly where, when and how it has established new breeding populations, and to assess the effect such populations have had on the existing biota. The mass of detailed information that Lever has compiled is clearly and systematically laid out, with separate world maps for each species to show its original range and the location of its naturalized populations, and an excellent line drawing by Robert Gillmor. Each entry also contains separate accounts for each new locality where a species has become established.

As described in Lever's introduction, birds have found their way to new parts of the world for a variety of reasons. Many were deliberately taken by colonists to provide food or sport, and some, notably songbirds, were taken simply to provide fond memories of home. More recently, birds have been transported for scientific reasons, such as in attempts to control pests or in efforts to establish new populations in the name of conservation when their original habitat is under threat. Some introductions are recent and precisely documented, others are ancient and of uncertain date. The common pheasant, for example, may well have been brought from the Far East to Greece by Jason and the Argonauts around 1300 BC, to be subsequently introduced to the rest of Europe by the Romans.

In many instances the long-term effect of introductions has been deleterious to indigenous birds. Alien species frequently carry diseases to which they have some immunity but to which the indigenous species do not: the effect of avian malaria on many of the endemic species of Hawaii has been devastating. Some species,

notably the house sparrow, starlings and pigeons, have become agricultural pests wherever they go, and where introduced species thrive and increase in numbers they generally do so at the expense of native birds. In America, for example, European starlings displace a variety of hole-nesting species from their nest sites. Moving birds from one part of the world to another can lead to hybridization between species that have not had the opportunity to evolve effective reproductive isolation. Thus, the mallard, exported several times from Europe, has hybridized with a number of duck species in Australia and New Zealand.

Competition from introduced species is only one of several factors that have contributed to the demise of the 75 or so species of bird that have become extinct since 1600. As detailed in Fuller's beautiful book, human greed, vanity or just blind stupidity have been the primary causes of the extermination of birds, for it is clear that, although extinction is the ultimate natural fate of all species, nearly all recently recorded extinctions among birds can be attributed to the direct or indirect effects of human civilization. Colonists from Western Europe have been the worst culprits in this sorry story.

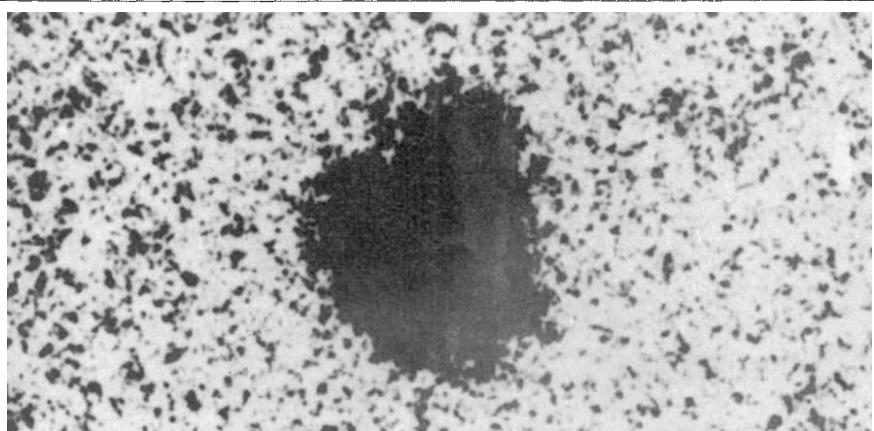
Fuller's book, as it is intended to be, is a worthy successor to Walter Rothschild's classic *Extinct Birds*, published in 1907. The text is the product of an impressive exercise in scholarship; the author has scoured obscure references, primarily to eye-witness accounts by travellers and explorers, and scrupulously avoids speculation or unsupported inference. Wherever possible, each species is illustrated by a colour painting executed by someone who had actually seen the bird in question. The reproductions are excellent and those by such notable artists as Audubon, Keulemans and Lear, although familiar to most ornithologists, add enormously to the attractiveness and scientific value of

the book. The book is arranged taxonomically and the length of each entry reflects the extent of our knowledge of each bird.

Although they deal with very different aspects of ornithology, these two books complement one another and have much in common. Both are superbly produced and the respective publishers are to be congratulated on doing justice to the labours of their authors. There is, however, one respect in which both books are of limited value. Although each is meticulously researched and comprehensive in its coverage of the material available, neither author provides a real synthesis of his subject matter. Lever's introduction gives a brief overview of the phenomenon of naturalization, its causes and its consequences, but does not set the phenomenon in the more general context of how the birds of the world are faring at the present time. The format of Fuller's book provides no space for an appraisal of the ultimate reasons why birds become extinct, or of the conservation measures that can be, and in some cases are being taken to prevent further extinctions.

Both books are essentially works of reference and, as such, will be invaluable for future authors who do wish to attempt a more thoughtful and analytical study of the factors which, in a rapidly changing world, cause some birds to flourish while many more decline. They present two sides of a single coin; as we human beings set about the destruction of our habitat, we eliminate rare, exotic and specialized birds while creating openings for a much smaller number of generalist, opportunistic species. The resulting homogeneity in the world's avifauna is a heavy price to pay for the reduction in diversity that we have brought about. □

Tim Halliday is Reader in Biology at The Open University, Walton Hall, Milton Keynes MK7 6AA, UK, and author of *Vanishing Birds, Their Natural History and Conservation* (Sidgwick & Jackson, 1978).



Mis-shapen identity — a mysterious bump first observed in 1978 on photographs of Pluto (here seen on the upper edge), which moves systematically round the planet and which turned out to be a satellite, Charon. The picture is taken from The Universe and Life by G. S. Kutter, which follows the evolution of matter from the big bang to Homo sapiens, with equal emphasis on both physical and biological evolution. Published by Jones & Bartlett, Boston, the book costs \$40.00.

How did they all get that way?

R.J.P. Williams

Silicon and Silicones: About Stone-age Tools, Antique Pottery, Modern Ceramics, Computers, Space Materials and How They All Got That Way. By Eugene G. Rochow. Springer-Verlag: 1987. Pp.181. Pbk DM 24.80, \$16.95.

EVERY now and then one comes across something different and what a blessing it is. This book is overtly about silicon, its chemistry and physics, and yet it is about something quite other. It is almost an autobiography.

Professor Rochow wants to educate us in a gentle way about silicones and he does this wonderfully well; any student could learn with pleasure from him about silicon, silicates, organosilanes and so on as they are described here. But the book is also about the excitement, challenge and fun of being a chemist and discovering chemicals. Here is a man who is not afraid to tell stories, quote poems, and laugh at

himself (and you). He literally loves silicones and treats his subject with reverence. While engaged in the description of this chemistry you are suddenly told how to make a 5MHz NMR spectrometer (for nothing) just in order to look at silicon NMR. Elsewhere the book meanders sweetly into fields of stone-age flint, semiconductors, insulators, glasses, ceramics — in fact anywhere the river silicon flows. "Sweet silicon run softly till you end your song."

At the end of the book Professor Rochow turns to silicon in biology. Only here does he miss a trick, for while he knows the inanimate silicon the lively variety is apparently less familiar to him. I'd like the chance to show him another glory of his love, the silicon of the radiolaria, of the nettle hair, of the protozoa.

I do not wish to close on this seeming criticism but rather to stress how this book reminded me of what has been lost — fun and enthusiasm for chemistry. Do read it and forget about granting agencies. □

R.J.P. Williams is Royal Society Napier Research Professor in the Department of Inorganic Chemistry, University of Oxford, and a Fellow of Wadham College, Oxford OX1 3PN, UK.

Hidden agenda

Tony Sudbery

Quantum Implications: Essays in Honour of David Bohm. Edited by B. J. Hiley and F. David Peat. Routledge & Kegan Paul: 1987. Pp.455. £25, \$49.95.

A STRIKING feature of the history of quantum mechanics is that among those who have done most to formulate the theory are those who have been the loudest in dissent from it. David Bohm is known to generations of students as the author of one of the clearest and most thoughtful textbooks of quantum mechanics, in which he gave the theory in its orthodox form (the 'Copenhagen interpretation') a reasoned and careful account which still, after nearly 40 years, serves as a standard. But he is also famous as the man who challenged the doctrine that alternatives to the Copenhagen interpretation were impossible, and his name is attached to a form of quantum mechanics (the 'causal' or 'hidden variables' version) which few students will learn in conventional courses. The interplay of conventional and unconventional is evident throughout this collection of essays, published to honour Professor Bohm on the occasion of his retirement from Birkbeck College.

The essays have been neatly arranged by Basil Hiley and David Peat to form a continuous fantasia on the themes that

have concerned Bohm throughout his life as a physicist. After an introduction in the form of an intellectual biography by the editors, Bohm himself describes the origin of his interest in hidden variables as a way of supplementing and making intelligible the usual version of quantum mechanics, and the development of this interest into his current concept of the 'implicate order'. We are then taken back to an early interest of Bohm's, many-particle quantum systems, in papers by his students Eugene Gross and David Pines on collective phenomena. Both are lucid and interesting, and accessible to the non-specialist.

There is a possible link between many-particle theory and the foundations of quantum mechanics, if A. J. Leggett is right in his theory that the puzzles of quantum mechanics will be resolved by the emergence of new laws which apply only to complex systems. Leggett describes this speculation in the course of a beautifully written article in which he reflects, with rare clarity, on the paradox that quantum mechanics, while purporting to describe phenomena consisting of discrete events, presents a theory in which events don't happen.

Links between quantum mechanics and macroscopic systems are also discussed by T. D. Clark, in a more technically detailed article than most on the construction of 'macroscopic quantum objects', and Roger Penrose, who speculates on the possible relevance of two apparently totally different topics, of gravity and

consciousness. Other papers on the foundations of quantum mechanics are contributed by Bernard d'Espagnat, John Bell, Richard Feynman, Geoffrey Chew and Henry Stapp. D'Espagnat's article is characteristically and judiciously philosophical, Feynman's characteristically and boldly pragmatic: let us, he suggests, simply allow that probabilities can sometimes be negative.

The papers by Bell on the one hand, and Chew and Stapp on the other, interestingly hint at a possible convergence of ideas on quantum mechanics. Although they are associated with rather different positions, they arrive here at what seem to me to be closely related formulations in their search for a conceptually satisfactory form of quantum field theory.

In Bohm's causal version of quantum mechanics, particles have definite trajectories in space and time, and their motion is controlled by a 'quantum potential' whose effects do not diminish with distance. This approach has been vigorously developed by groups led by Bohm and Hiley at Birkbeck and J.-P. Vigiér in Paris. Vigiér and some of his co-workers contribute a long article surveying their recent work in this area. To many physicists this work is reminiscent of nineteenth-century work on the structure of the ether, and there is indeed a paper by P. R. Holland and C. Philippidis on electromagnetic potentials as dislocations in a relativistic ether.

Conceptual problems in quantum mechanics are closely related to problems concerning time and irreversibility; such problems are the subject of papers by Ilya Prigogine and Yves Elskens, and by Y. Aharonov and D. Albert. David Finkelstein muses in a delightfully elegant and stimulating way on time, space and physics, and Robert Rosen contributes an excellent essay, cogent and clear-headed, on the relation between physics and biology.

In recent years Bohm has been concerned with his concept of 'the implicate order', a view of quantum physics based on the metaphors of the hologram and of ink drops in glycerine. The notion of implicate order is brought to bear here on neurology, psychiatry, meditation and aesthetics, and is finally the subject of a conversation between Bohm and Renée Weber. The meaning and significance of this part of the book remain, to this reader at least, appropriately implicate.

Quantum Implications is a fitting tribute to one of the most searching thinkers in modern physics, and will become a standard reference work on the concepts of quantum mechanics. □

Tony Sudbery is a Senior Lecturer in the Department of Mathematics, University of York, York YO1 5DD, UK, and author of Quantum Mechanics and the Particles of Nature (Cambridge University Press, 1986).

Eighteen eighty eight and all that

W. F. Bynum and J. L. Heilbron

This year the pneumatic tyre will have been around for 100 years, while NASA reaches the age of 30. Those wishing to join the Anniversary Waltz should read on.

LEST last year's great harvest of anniversaries, led by the third centenary of the publication of Newton's *Principia*, leave the impression that history and historians have exhausted themselves, we notice here a few events in need of celebration this year. We recommend that the festivities should begin early. For in 1888 Jupiter and Saturn move their meetings from one set of zodiacal signs to another, as they do every 200 years by the reckoning of the old astrologers. The change always presages a disaster — in 1788, the French Revolution, in 1588, the destruction of the Spanish Armada, to go back only two cycles.

Among anniversaries the hundredth seems to have the strongest appeal, perhaps because it refers to a past remote enough to be picturesque yet close enough to be familiar, like the early life of one's grandfather. There is nothing sacred about multiples of ten, twenty-five or one hundred, however. An exhibition was contemplated four years ago to commemorate the 158th anniversary of the introduction of the giraffe into France; the surprising thing perhaps is that this happy event did not occur until 1826, two decades after J.B. Lamarck had made that animal's neck a *cause célèbre* in biology. For the most part, though, we have chosen round numbers as occasions to reflect on then and now, to understand why a particular historical path was taken, or merely to be grateful for some person, institution, discovery, publication or invention. We begin our enumeration here with hundredth anniversaries, which we take as our unit; we proceed to multiples, and end with rational fractions.

1888 (1.0 centenary)

The celebration of anniversaries is scarcely possible without an instrument invented and first marketed 100 years ago. George Eastman's Kodak camera (the name was part of the invention) brought photography within the reach of anyone able to press a button and pay \$10 for the processing and printing of the exposed film (each roll had 100 exposures). The

French physiologist Etienne-Jules Marey devised the first modern cine camera in the same year. Consequent opportunities in the mass market have made the photographic industry one of the most progressive and research-intensive of the second industrial revolution.

By 1888 photography was already an important tool in scientific inquiry, par-

C.W. Tombaugh in 1930: $(1930 + 1846) = 2 \cdot 1888$.

No doubt Fritjof Nansen would have been happy to carry a Kodak when he skied across Greenland in the summer and autumn of 1888. Nonetheless the readers of his scientific adventure, *First Crossing of Greenland*, were willing to accept his statement, which conflicted with the most

Kodak Museum authoritative conjectures, that the island is covered with glaciers. The Greenland crossing established Nansen as an explorer and as a meteorologist. He had ample time to study the island because he and his party returned to the coast too late for the last ship and spent the winter with the Eskimos. He would have found little use for John Boyd Dunlop's pneumatic tyre, patented in 1888, but John Laud's ball-point pen of the same year might have served its turn. Curiously, Laud's invention did not catch on, and exactly half a century later the Hungarian brothers George and Ladislau J. Biro re-invented the writing instrument which has acquired their name.

No sooner have the meetings honouring the centennial of the National Institutes of Health in Bethesda, Maryland, finished, than attention moves to another

great medical research establishment, the Institut Pasteur. Indeed, the official birthday began in late 1987, presumably to commemorate the fund raising, although the Institut's doors did not open until the autumn of 1888, a tangible expression of the hope which Louis Pasteur's research had kindled in men and women throughout the world. Over the years it has nurtured its fair share of Nobel laureates, from Élie Metchnikoff and Jules Bordet to Jacques Monod, André Lwoff and François Jacob. All French Nobel prizes for physiology or medicine have been won by individuals associated at some stage of their career with the parent or a satellite Institut Pasteur. The Duchess of Windsor's recent legacy should ensure the Institut's fiscal solvency for some time to come, and AIDS now occupies a position comparable to that of rabies at its birth. (Incidentally, the Nobel Committee may wish

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

In the round — George Eastman with the Number Two Kodak camera, photographed in 1890 two years after the Kodak's advent.

ticularly in astronomy. That year Hermann Carl Vogel, director of the then new Potsdam Observatory and one of the first astronomers to make systematic use of photography, obtained on film the first evidence of the radial velocity of stars by Doppler shifts in their spectra. At about the same time similar evidence allowed him to discover the existence of eclipsing binaries. The new *paparazzi* of the heavens obtained revealing pictures of many celestial objects in the 1880s, of which a three-hour exposure of the great Andromeda nebula, taken on 1 October 1888 and showing its spiral structure, deserves pride of place.

A subtler astronomical centennial, which we offer for its intrinsic interest and as a contribution to anniversarial method, is that of the average date of discovery of the outermost planets. Neptune was first observed by J.C. Galle in 1846, Pluto by

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Seeing stars — the Lick Observatory, one hundred years old in 1988, and a long-exposure picture of the great Andromeda nebula taken in October 1888 by Isaac Roberts.

to note a pregnant centenary in its own history: the death of Ascano Sobrero, the original discoverer of nitroglycerin.)

Microbiologists have more than just the Institut Pasteur to remember. In 1888 Gram's stain was four years old, but Hans C.J. Gram himself died exactly half a century ago. *Salmonella enteritidis* was isolated by August A.H. Gaertner in 1888, the year of Selman A. Waksman's birth. The Petri dish became a centenarian last year.

Drosophila melanogaster is a slightly later addition to the biological research scene, but its own laboratory life has been intertwined with that of another major research establishment eligible for a presidential telegram, the Marine Biological Laboratory at Woods Hole, Massachusetts. Designed to replicate the best features of the biological station at Naples, the MBL has never abandoned the 'marine' of its title, but its early workers, such as C.O. Whitman, E.B. Wilson, T.H. Morgan and Ross Harrison, ensured that fundamental problems of heredity, genetics and embryology were also part of its brief. It was established the same year as Theodor Boveri published the second of his classical studies on chromosome individuality in the cells of *Ascaris*. Boveri was undoubtedly aware that 1888 was the 0.5 centenary of Jacob Schleiden's generalization of the cell as the building-block of all plants. In 1988, the Australasian Association for the Advancement of Science is exactly half as old as the penal settlement at Botany Bay.

Physicists have a centennial perfectly suited to follow that of the Michelson-Morley experiment celebrated last year. In 1888 the odd effect of one spark on another, noticed by Heinrich Hertz during the researches that led him to the discovery of electromagnetic waves, was generalized by several physicists into the photoeffect. Those responsible —

Wilhelm Hallwachs, Augusto Righi and Alexandr Stoletow — established that ultraviolet light can discharge a negatively electrified insulated metal plate and give a positive charge to a neutral one. A decade later J.J. Thomson identified the charge carriers driven from the plate by the light as electrons. No more than the null result of the Michelson-Morley experiment, however, were the quantitative details of the photoeffect plausibly explained by the physical theories available around 1900. Relativity was needed to handle the first, quantum theory the second.

Anyone wishing to fête thermodynamics will have an unusual opportunity this year. In August 1888 Rudolf Clausius, who had found the principle of entropy, died in Bonn, a few weeks after his soul had flitted to Upper Silesia, into the newborn Otto Stern, who was to choose physics for a career under the inspiration of Clausius's writings. Stern began as a theorist specializing in thermodynamics. He later switched to experiments with atomic beams, for which he won the Nobel prize in physics for 1943.

Another Nobel laureate born in 1888 holds an interest not only for physicists and chemists, but also for numerologists and sociologists. He was Chandrasekhara Venkata Raman, who detected the quantum optical effect named after him in 1928, giving us both a 1.0 and a 0.6 centenary. Raman learned his physics in India, had no period of study abroad, and spent ten years as an accountant in the Indian civil service before becoming professor of physics at the University of Calcutta. He was the first Indian to be nominated for, and the first Indian to receive, a Nobel prize in either physics or chemistry.

An event of wide appeal is the hundredth anniversary of the celebration of the eighth centennial of the founding of the University of Bologna. The way the founding date of the University was estab-

lished deserves admiration. By 1886, Bologna had seen several universities far junior to its ancient Studio — Edinburgh, Heidelberg and Uppsala, for example — commemorate their origins. A bash was planned in Bologna for 1888, to inaugurate a monument to Vittorio Emanuele II. The founding of the University was accordingly fixed at 1088, and its celebration joined to the monument's. Bologna's seniority was conceded by the leading universities of Europe; its 900th anniversary this year, however artificial, commemorates not a single institution, but the endurance of universities in general and the ideals of free inquiry and critical learning they represent. The best-known scientist produced at Bologna during its 900 years was Luigi Galvani (1737–1798), who as anatomist, physiologist, chemist, physicist, physician and philosopher was a university in himself. He also qualifies for a centenary — the 2.5 of his birth — to within an accuracy of one part in 250.

1788 (2.0 centenary)

According to our calculations, double centennials fall next year to the great calculators of the *ancien régime*, Joseph Louis Lagrange and Pierre Simon de Laplace. In 1788 Lagrange gave the world his *Mécanique analytique*, a generalization of Newtonian mechanics without the use of pictures; and Laplace provided, in several enormous memoirs, proof that the Solar System is stable under the mutual gravitational interactions of its members. He put two memorable glosses on this result. The first, written on the eve of the French Revolution, makes a useful warning to those inclined to make analogies: "The stability in the system of the world . . . exhibits in the heavens the same intention to maintain order in the universe that nature has so admirably observed on earth for the sake of preserving individuals and perpetuating species". The second gave notice to agents of a providential God that Laplace had no need of that hypothesis.

Whether Laplace's countryman Georges Buffon, who died in 1788, needed God is still a matter of historical debate. He was said to have talked matter to the *Philosophes* and spirit to the Sorbonne. Certainly his great rival, Carl Linnaeus, needed the Deity and was convinced that the need was mutual. This Second Adam, as he liked to think of himself, named us *Homo sapiens* and much else besides. Although the 2.81 centenary of his birth may not be widely noticed, his vast collections were acquired by the English physician and botanist Sir James Edward Smith, who in 1788 founded a society to perpetuate the Swede's name. Established at a time when science in Britain was primarily the amateur's domain, the Linnean Society provided the

forum, 1.3 centuries ago, for the communication, by Charles Lyell and Joseph Hooker, of the Darwin–Wallace discoveries of natural selection. Ardent Darwinists will probably also note that the Great Amateur had first discussed the principle of natural selection in his private notebooks 0.2 centuries previously.

The eve of the French Revolution also witnessed its share of births and deaths. Among the former we may record the chemist Leopold Gmelin, the physicist Augustin Jean Fresnel and the phrenologist George Combe. Last rites were said for the eccentric physician John Brown, prematurely worn out by his favourite remedies opium and alcohol, the electrician Benjamin Wilson and the geologist John Whitehurst. Of geological note, too, was the publication of the article by James Hutton which grew into the two volumes of his *Theory of the Earth* seven years later. Not to be outdone by Laplace, Hutton also envisaged a universe of perfect order and stability, exhibiting, at least as far as the eye could see, “no vestige of a beginning, — no prospect of an end”.

1688 (3.0 centenary)

Persevering mathematicians were still hard at work reading Newton's *Principia*. Mrs Swedenborg, gravid with her son Emanuel, was otherwise engaged, as was Professor Swedenborg, her husband, whose subject was theology. Emanuel studied Newton on a trip to England and later wrote his own *Principia*: but ultimately he chose his father's path and sought the essence of things via an eight-volume commentary on the five books of Moses.

1588 (4.0 centenary)

The great comet of 1588 announced the births of Thomas Hobbes, Marin Mersenne and Étienne Pascal, and the beginning of the end of the Spanish Empire. The same year Tycho Brahe published his observations on the comet of 1577, to which he added a new geometry of the Universe, in which all planets circle the Sun and the Sun, with its brood, circles the Earth. This Tyconic system enjoyed great popularity among fence sitters until Galileo — who in 1588 was lecturing on the geometry of Dante's *Inferno* — omitted it from his discourse on the chief world systems. One reason that conservative folks liked Tycho's geometry was that it did not require them to perplex themselves with the physics of a spinning Earth. The volume in which Galileo presented in detail the theory of motion that helped many people advance from Tycho to Copernicus, the *Two New Sciences* of 1638, will have a 3.5 centenary this year. So, too, will William Gascoigne's micrometer. There was no comet in 1638 to announce the arrivals of Niels Stensen or Frederick Ruysch.

1488 (5.0 centenary)

Bartolomeu Diaz returned to Portugal after turning the Cape of Good Hope and opening the possibility of an eastern sea route to China. Christopher Columbus also returned to Portugal, from Spain, where he had not found a backer for his plan to reach the Orient by sailing west. That was not because educated people then doubted that the world was round, but because Columbus, by compounding all the errors of measurement he could find, made the world much smaller than the learned thought likely. As everyone knows, Columbus had to wait for his grant. Wise centenarists are already asking for theirs for the celebrations of the ill-advised investment of Isabella of Spain that will occur throughout 1992.

1288 (7.0 centenary)

Our oldest anniversarian also has the longest name (eleven words long), but is known to the world as Ibn al-Nafis. He died in Cairo in 1288. Some of his prolific writings remain unpublished, but Ibn al-Nafis more than once enunciated a clear description of the pulmonary circulation. More ink will undoubtedly be used trying to prove that Michael Servetus knew of his work three centuries later: the verdict will probably always be the proverbial Scottish one. Meanwhile, Europeans in 1288 were less interested in the circulation of blood than in the circulation of currency: share dealings were first recorded in Stora Kopparbergs, in what may be the world's oldest company.

1913 (0.75 centenary)

From the remote we return to the recent, although the First World War is seen by many as an event which demarcates us from *temps perdu*. The 0.75 centenary is roughly an average human lifespan now. Appropriately, 1913 was not a bad year for medical research. Bella Schick developed the test for diphtheria immunity, and Elmer McCollum and Marguerite Davis isolated vitamin A. The Medical Research Committee (now the Medical Research Council — MRC) was also born. It is not certain whose brainchild it was, although the civil servant Sir Robert Morant has the best claim. The MRC was the almost incidental by-product of the 1911 National Health Insurance Act, supported by one penny from each person covered under the Act. Tuberculosis was its original brief, but by the time the Committee was set up a more broadly based research front was planned. The outbreak of the First World War influenced the Committee's early history, just as the Second was to delay the move to its primary research headquarters in Mill Hill, north London. One of the MRC's more visible outposts, the Laboratory for Molecular Biology in Cambridge, celebrated a 0.4 centenary last year. James

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Looking out (top) Heinrich Hertz, who a century ago was the first to observe the photoeffect; and looking in, Luigi Galvani, most famous son of the University of Bologna which celebrates its 900th anniversary this year.

Watson and Francis Crick (working in Cambridge) and Maurice Wilkins have a 0.35 centenary for their papers on the structure of DNA, published in this journal.

A few may need reminding that in 1913 Frederick Soddy coined the term ‘isotope’, Niels Bohr announced his quantized atom and Johannes Stark split the spectral lines in an electric field. The International Health Board of the Rockefeller Foundation was organized and Henry Ford was pioneering his methods of automobile assembly through conveyor belts. Charles Dawson published the first paper on Piltdown Man.

1938 (0.5 centenary)

We should be grateful that the human race is still around to notice the most influential discovery in physical science made in 1938. This was the realization, by Otto Hahn and Fritz Strassmann, that the radioelements obtained by irradiating uranium

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

October 1954, and the transistor trio (left to right: John Bardeen, William Shockley and Walter Brattain) pose for the press in the knowledge they are strong contenders for the Nobel prize. Their number in fact came up two years later.

by neutrons that they had been studying were not heavier than uranium, as most knowledgeable investigators thought, but considerably lighter. The explanation of their observations by the concept of nuclear fission, proposed by their former collaborator Lise Meitner, who had been driven out of Germany by the Nazis, and Otto Robert Frisch, started the investigations that brought atomic bombs and nuclear power. The episode contains several lesser centennials: Hahn and Meitner both died in 1968 (0.2 centenary); the Kaiser-Wilhelm-Institut für Chemie, in which Hahn and Strassmann worked, opened just before 1913 (0.75+ centenary); and most of the uranium with which they worked had an isotopic weight of 238 (17.5 centenary). And that is not nearly all. Martin Heinrich Klaproth first identified uranium 200–1 years ago (a 2.0–0.5% centenary); and he named it as he did out of fellow feeling for the discoverer of the planet Uranus, Wilhelm (William) Herschel, who was born in 1738 (a 2.5 centenary).

1928 and 1948 (0.5 ± 0.1 centenary)

The year 1928 is full of significance for the student of electrons. Almost simultaneously that year H.A. Lorentz, the grand master and main creator of the classical theory of electrons, died, and P.A.M. Dirac gave birth to the new theory, the relativistic quantum theory of electrons, one of the all-time triumphs of mathematical physics. Enthusiasts freer with chronology than we might add to the package the demonstrations of the wave properties of the electron by C.J. Davisson, L.H. Germer and G.P. Thomson, which were

published in 1927. J.L. Baird's transatlantic television transmission and Alexander Fleming's paper on penicillin were other features of 1928.

Nineteen forty eight marked a watershed for accelerator physics. Until then the only source of mesons was God, who provided them stingily in cosmic radiation. In 1947 Cecil Powell caught one sort of meson (which he named 'pi') in the act of transformation into the common cosmic-ray meson ('mu'). Early in 1948, Eugene Gardner and G.C.M. Lattes detected pi mesons made in abundance by man in the then new synchro-cyclotron at Berkeley. Thereafter machines have made elementary particles and Nobel prizes in profusion.

In the summer of the same year, newspapers announced that the world had been enriched by a device scheduled to revolutionize the electronics industry. This was the transistor, invented in the Bell Telephone Laboratories by John Bardeen, Walter Brattain and William Shockley. The importance of the transistor for everything from stereos to Star Wars needs no advertisement here. What may deserve notice is that Bell Laboratories held back notice of the discovery not only for proprietary reasons, but also to make room for a security check even though the government had not supported the research. The military came, saw and did not classify.

Spare a thought, too, for the National Health Service, which came into being on the Appointed Day, 5 July 1948.

1918 and 1958 (0.5 ± 0.2 centenary)

In 1918 "the greatest chance we ever

had to advance research in America", as George Ellery Hale, director of the Mount Wilson Observatory, characterized the war to end all wars, came to an end. Among the attractive research projects awaiting demobilized scientists was the test of general relativity. The first in the field was a team under William Wallace Campbell, head of the Lick Observatory, which opened its doors in 1888 (a 1.0 centenary). Campbell's group deduced from its observations of the total solar eclipse of June 1918 that the Universe was on their side (the clouds parted less than a minute before totality) and not on Einstein's. It might be best to postpone celebration for a year, when the British expeditions to observe the eclipse of May 1919, which came closer to confirming the predictions of general relativity, have an anniversary.

The year 1958 saw the United States react to the launch of Sputnik the year before by setting up the National Aeronautics and Space Administration. The progress of NASA in overtaking the Soviet Union in space did not satisfy the new president, John F. Kennedy. In May 1961, shortly after taking office, Kennedy proposed that the nation should commit itself to putting a man on the Moon and returning him to Earth before the decade was out. Had all gone as his optimistic advisers anticipated, the man would have landed in 1968 (a 0.2 centenary); but Neil Armstrong did not take his step for mankind until July 1969. The Apollo project has capital importance from many points of view, but perhaps primarily as a reaffirmation of the heady lesson of the Manhattan project — when the physics and engineering are ripe and money is plentiful, *Homo faber* can do anything, except ensure the survival of *Homo sapiens*.

We recognize that what we have mentioned are but a few morsels plucked almost at random from the groaning table of eligible anniversaries. Please do not send us additions or corrections unless remaining silent would injure your health. We have now turned to anniversaries base twelve, with $1.0.0 = 144$ as our 'centenary'. Some preliminary results: at one-quarter unit, the first thermonuclear device; at one-half unit, the first systematic treatments of general relativity by Einstein and of the quantum theory of the atom by Arnold Sommerfeld; at one unit, James Joule's demonstrations of the mechanical equivalent of heat; and at two units, 1700, the first year *anno Domini* visible by four not counted as a leap year. The method appears at least as promising as the vulgar reckoning. □

W.F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BP, UK. J.L. Heilbron is in the Office of History of Science and Technology, University of California at Berkeley, California 94720, USA.

Action principles in nature

John D. Barrow* & Frank J. Tipler†

* Astronomy Centre, University of Sussex, Brighton BN1 9QH, UK

† Department of Mathematics and Department of Physics, Tulane University, New Orleans, Louisiana 70118, USA

Physical theories have their most fundamental expression as action integrals. This suggests that the total action of the Universe is the most fundamental physical quantity, and hence finite. In this article it is argued that finite universal action implies that the Universe is spatially closed. Further, the possible spatial topologies, the types of matter that can dominate the early universe dynamics, and the form of any quadratic additions to the lagrangian of general relativity are constrained. Initial and final cosmological curvature singularities are required to avoid a universal action singularity.

IN most discussions, the initial cosmological singularity predicted by classical general relativity is regarded as an indication of the breakdown of the theory rather than a prediction of an extant singularity on the boundary of space-time. Indeed, the prediction of a past cosmological singularity may be regarded as insecure because of the breakdown of general relativity at high density or a violation of the energy conditions assumed in the proof of the singularity theorems. However, since we have no tested theory of quantum gravitation to supersede general relativity nor any observational evidence for the existence of matter fields which violate the strong or weak energy conditions, the question of whether cosmological singularities do in fact arise is still an open one. In this article we demonstrate that the removal of the local 'infinity' created by the existence of cosmological curvature singularities introduces a singularity in a possibly more fundamental global quantity, the total action of the Universe. We show this action singularity in general cannot be removed by a violation of the energy conditions or known quantum gravity effects. We shall also investigate what properties are required of the Universe for the total action to be finite. We shall find that a number of interesting constraints are placed upon the structure and content of the Universe by this restriction.

Finite action in classical and quantum fields

Max Planck¹ was perhaps the first in the twentieth century (see ref. 2 for a discussion of earlier work) to argue strongly that the action is the most basic quantity in physics. He, and later Dirac³, justified this on the ground of the manifest Lorentz (and $GL(4, \mathbf{R})$) invariance of the action integral. In contemporary physics, the action is used as the starting point of fundamental theories because of its invariance under gauge transformations of Yang-Mills and supersymmetric fields, and also because of the central importance of the path-integral method of quantization, which uses the action in a central way^{4,5}.

It is well known that the requirement of finite action has important physical significance for actions defined in Minkowski and euclidean space. Recall that a finite-action solution to the Yang-Mills-Higgs equations for the fields (A, Φ) in Minkowski space is called a soliton⁶ and, if the space is euclidean, a finite-action solution to pure Yang-Mills ($\Phi = 0$) is termed an instanton⁷. These actions are related, as in general the energy of the static configuration (one on which the fields (A, Φ) are time-independent) of the Minkowski action equals the euclidean action in one dimension lower⁶.

In euclidean space, the finite-action requirement imposes many strong restrictions on the solutions to the Yang-Mills-Higgs equations⁹; for example: (1) the fields must die off

sufficiently rapidly at euclidean infinity; (2) the minima of the pure Yang-Mills action must be instantons; (3) instantons must be self-dual or anti-self-dual; (4) if $d > 4$, where d is the dimension of the euclidean space, then there are no nontrivial solutions to the Yang-Mills-Higgs equations; and (5) if $d < 4$, there are no nontrivial solutions to the pure Yang-Mills equations⁶.

The reason for considering $d = 4$ instantons—euclidean actions—is that the path integral is usually defined by Wick-rotating the Minkowski space action, S , to a $d = 4$ euclidean action, so that the path-integral weight-factor $\exp(-S)$ defines a Wiener measure. The classical euclidean solutions with finite action provide a starting point for a semi-classical approximation to the euclidean path integral, which can be analytically continued back into Minkowski space to obtain the physical path integral.

Of fundamental importance is the fact that instantons are useful only in the semi-classical approximation; in euclidean space, the paths with finite action are a set of measure zero^{6,8} in the function space of paths, in spite of the weight $\exp(-S)$ being zero for infinite action. As is clear from the proof of this for the simple harmonic oscillator⁸, the infinite action paths will have an overwhelming importance in any natural measure on euclidean space. Euclidean infinity will dominate the path integral and make its presence felt in any local experiment. Thus if the finite-action principle is to apply in quantum as well as classical mechanics, the function space upon which the path integral of the entire Universe is defined must be such as to give dominant weight to the finite action paths, or (what is physically equivalent), the universal action operator and possibly other observables (self-adjoint operators) must be bounded operators. This bounded operator requirement is enormously restrictive. We shall analyse its implications elsewhere.

As discussed most thoroughly by DeWitt⁹, if we regard the action merely as a device to obtain the dynamical equations $\delta S / \delta \Phi_i = 0$, where Φ_i are the dynamical fields, then there are many different actions leading to the same physics. Nevertheless the finite-action principle has physical content because the functional derivative $\delta S / \delta \Phi_i$ makes sense only if S is a finite functional. It is to ensure finiteness that the action integral is often restricted to a bounded domain in an unbounded space (with appropriate stationary boundary conditions) when its functional derivative is evaluated.

Finite action in Friedman universes

We now consider the implications of finite action in cosmology. The general action¹⁰⁻¹² for Einstein's gravitational theory has

three pieces ($8\pi G = 1 = c$). Formally

$$S \equiv S_g + S_m + \text{boundary term};$$

that is:

$$S = \frac{1}{2} \int_M (R + 2\Lambda) \sqrt{-g} d^4x + \int_M (L_m \sqrt{-g}) d^4x + \int_{\partial M} (\text{tr } K)(\sqrt{\pm h}) d^3x \quad (1)$$

where R is the 4-d Ricci scalar, Λ is the cosmological constant, L_m is the matter lagrangian (defined as in ref. 13) and S_g and S_m are the gravitational and matter actions respectively. The integration is over a 4-manifold M with space-time metric g . The boundary of M is ∂M , where ∂M has induced metric h_{ab} and extrinsic curvature K_{ab} . The plus (minus) sign in $(\pm h)^{1/2}$ is chosen if the boundary is spacelike (timelike) respectively. The boundary term ensures that the metric g satisfies the Einstein equations when S is an extremum under all variations of g which vanish on the boundary. If S is to be the universal action, the appropriate 4-manifold over which the integration is performed must be the whole of space-time, so there is no boundary. Therefore, unless otherwise stated, we shall henceforth drop the boundary terms from equation (1). The fact that the ugly boundary terms can be removed from equation (1) only if the universal action is finite is itself an argument for the principle of finite action. The universal action will be singularity-free (that is, finite) only if each term in equation (1) is finite.

It is possible of course that S_g and S_m could separately diverge while the sum remains finite. If the Einstein equations with $\Lambda = 0$ hold, then $S_g = -S_m$ whenever $T_a^a = L_m$, so in this case the total action will cancel to zero. But in general $T_a^a \neq L_m$. For example, in the perfect fluid case with equation of state $p = (\gamma - 1)\mu$, $L_m = -2\mu$, and $T_a^a = (3\gamma - 4)\mu$, so $T_a^a = L_m$ only when $\gamma = 2/3$. Except when $\gamma = 2/3$, $S_g + S_m$ will be finite only if both S_g and S_m are finite. The same is true for general electromagnetic fields, for which $L_m = -(8\pi)^{-1} F_{mn} F^{mn} = -(4\pi)^{-1} (E^2 - B^2)$ and $T_a^a = 0$. Thus, unless $E = B$, we will have $L_m \neq T$, and we would generally have $E = B$ only for pure radiation fields.

For massless scalar fields, $L_m = -\Phi_{,m} \Phi^{,m} = T_m^m$, so any separate divergences in S_g and S_m will cancel in the sum along the physical trajectory. Nevertheless, even in this case it is reasonable to require that S_g and S_m are separately finite. For all the physics is contained in the statement that the total action is stationary under variation of all physical fields. Since the matter fields are present only in S_m , a variation of $S_g + S_m$ will make sense only if S_m is itself a finite quantity near the physical trajectory. As a general rule, the finiteness of one term in equation (1) will imply the finiteness of the others because of the field equations, so it will generally (but not always) be sufficient to establish the finiteness of one term in equation (1) to establish the finiteness of all terms. Thus in what follows we shall assume that S_g and S_m are separately finite.

If the Universe is roughly homogeneous and isotropic (that is, if the Cosmological Principle holds) then it must be spatially closed, or else the integration over space by itself would cause all terms in equation (1) to diverge (assuming that the lagrangians are not identically zero). In particular, the steady-state theory, both in the original version developed by Bondi, Gold, Hoyle and Narlikar¹⁴ (derived from the Pryce action), and also in the reincarnation now defended by Gott¹⁵, Guth¹⁶ and Linde¹⁷, has singular universal action, because the steady-state theory in either form is spatially infinite.

But spatial closure is not sufficient for finite action. The action of the static Einstein universe is also infinite because $L_m \sqrt{-g}$, $R \sqrt{-g}$ and $\Lambda \sqrt{-g}$ are independent of time, and so the integrals in equation (1) diverge when the time integration is carried out.

In both the Einstein static universe and the steady-state universe, the absence of singularities in curvature invariants is purchased at the price of a singularity in the universal action. This is a general phenomenon.

To see this, consider all spatially closed Friedman universes containing a perfect fluid. We have that $R = -T_a^a \propto \mu - 3p = (\gamma - 4/3)\mu$. In all Friedman models, $\mu \propto a^{-3\gamma}$. In the $k=0$ models, $a(t) \propto t^{2/3\gamma}$, ($\gamma \neq 0$), where t is the proper time. This proportionality will also hold as $t \rightarrow 0$ in the $k \neq 0$ models (provided $\gamma > 2/3$). Thus, near each singularity at $t=0$, we have $\int R \sqrt{-g} d^4x \propto (\gamma - 4/3) t^{(2-\gamma)/\gamma}$ and $\int L_m \sqrt{-g} d^4x \propto t^{(2-\gamma)/\gamma}$, provided $\gamma \neq 2$ or 0 . If $\gamma = 2$, the $t^{(2-\gamma)/\gamma}$ factor is replaced by $\ln t$. Hence, when $\gamma \geq 2$, the action diverges as $t \rightarrow 0$. (One can show that the action for the full $k=+1$, $\gamma=2$ solution diverges, rather than just that of its asymptote as shown here.) The $\gamma=0$ case is equivalent to a vacuum universe with $\Lambda \neq 0$ that must be either de Sitter or anti-de Sitter space-time¹³, and we pointed out above that the corresponding contribution to the action in equation (1) diverges because of the infinite range of t . Similarly, the action for all $k=0$ spatially-compact models (constructed by topological identifications of $\mathbf{R}^3$) diverges for all γ because the $t^{(2-\gamma)/\gamma}$ factor will either diverge as $t \rightarrow 0$ (if $\gamma < 0$ or $\gamma \geq 2$) or as $t \rightarrow \infty$ (if $0 \leq \gamma \leq 2$). The curvature term in the Friedman equation is k/a^2 , and it will dominate the $\mu \propto a^{-3\gamma}$ term for $a \rightarrow \infty$ whenever $\gamma > 2/3$. When $\gamma > 2/3$ and $k=-1$, $a \propto t$ for large t , so $S_g \propto t^{(4-3\gamma)}$, which diverges as $t \rightarrow \infty$ for all $\gamma < 4/3$. In fact, the action for the $k=-1$ spatially compact Friedman models diverges for all γ because the boundary term in the action (1) diverges at least as fast as t^2 as $t \rightarrow \infty$.

Thus, a perfect fluid Friedman model can have a finite action only if $k=+1$. All $k=+1$ and $\gamma \leq 2/3$ Friedman models expand forever with asymptotic behaviour approaching that of the $k=0$ model as $a \rightarrow \infty$. Hence, as shown above for the $k=0$ case, S_g must diverge as $t \rightarrow \infty$ for $0 \leq \gamma \leq 2/3$ and diverge as $t \rightarrow 0$ for $\gamma < 0$. Furthermore, if $\gamma > 2/3$, all $k=+1$ Friedman models begin and end with singularities^{18,19}; that is, the time integration is over a finite range, so if $2/3 < \gamma < 2$, S_g and S_m will both be finite because we showed, above, that they are finite near each singularity for $2/3 < \gamma < 2$. If ∂M consists of two ($k=+1$) slices of constant time, then on a slice near a singularity the boundary term in equation (1) is proportional to $(2/\gamma) t^{(2-\gamma)/\gamma} \int d^3x$, which vanishes as $t \rightarrow 0$ for $0 < \gamma < 2$. Thus the boundary term vanishes whenever S_m and S_g are finite.

In summary, the action of a Friedman universe containing a single perfect fluid is finite if and only if $k=+1$, and $2/3 < \gamma < 2$.

If there are several fluids, the limits on γ can change. For example, for the radiation-filled closed Friedman models with $\Lambda > 0$, we have $a^2(t) = 1 - \cosh 2t + \beta \sinh 2t$, where β depends on the ratio between μa^4 and Λ . The action will be finite only for β in the range $(0, 1)$.

If we have several fluids, with one dominating for small and another for large $a(t)$, then similar arguments to those given above show that finite action is consistent with infinite past or future time evolution of the universe only if $k=0$, $a \rightarrow \infty$ as $t \rightarrow \infty$, and a fluid with $\gamma > 2$ dominates as $a \rightarrow \infty$. But this case is unphysical since the action will diverge if an ever-expanding universe contains a radiation field or a single stable massive particle. Furthermore, it is unphysical since $\gamma > 2$ fluids permit acausal propagation of sound waves.

Thus we see that in all cases of physical interest the finiteness of the action results from the fact that the range of the time integration is finite. This finite range in turn is a manifestation of the fact that the closed Friedman universe begins and ends in singularities. In other words, the finiteness of the action depends on the existence of curvature singularities a finite temporal distance in the past and future. Without these curvature singularities, the time evolution would proceed without limit, and this infinity would manifest itself in infinite action. Thus in general, there is a trade-off between space-time singularities: a singularity in the action is avoided only at the price of a singular-

ity in curvature invariants, and *vice versa*. In cosmology, some sort of singularity seems inevitable. Even an ultimate 'Theory of Everything' (TOE) would in general eliminate the classical singularity only at the price of a singular universal action. For if we ignore the possibility of cyclic time because of the second law of thermodynamics, the elimination of the classical singularity in general would imply an infinite 4-volume in which the classical action (1) is an excellent approximation to the ultimate quantum TOE action²⁰.

The finite-action principle in cosmology restricts the allowed forms of matter just as it does in euclidean Yang-Mills theory. Scalar fields in closed universes are also ruled out as the dominant matter source near a singularity, as one would expect since scalar fields behave like $\gamma=2$ perfect fluids in the $a \rightarrow 0$ limit. For the massless scalar field, $\Phi(t)$, in the Friedman universe ($L_m \propto (d\Phi/dt)^2$), near each singularity we have $d\Phi/dt \propto t^{-1}$ and $a(t) \propto t^{1/3}$ (this is exact in flat models), so that on approach to each singularity the matter term in the action goes as $S_m \propto \int (d\Phi/dt)^2 a^3 dt \propto \ln(t)$, which diverges as $t \rightarrow 0$. Massive scalar fields also give a divergent S_m , since the mass term is either negligible near the singularity, or causes a bounce²¹.

The divergence of the scalar field action near singularities suggests that scalar fields are inappropriate matter sources to test the strong cosmic censorship hypothesis²², as has recently been done by Christodoulou²³. The divergence of the action integral is as bad as the shell-crossing singularities generated by collapse of shells of dust²⁴. Just as shell-crossing singularities can be removed by a slight change in the equation of state, so we would predict that the violation of cosmic censorship caused by the divergence of scalar fields outside event horizons can be removed by modifying the equation of state to give an action integral that is bounded near a singularity. Further evidence for our prediction is the fact that static, spherically symmetric space-times containing massless scalar fields possess no event horizons, but instead have naked singularities. Analogously, vacuum static, spherically symmetric, Brans-Dicke solutions have no event horizons, just naked singularities, unless the Brans-Dicke scalar field is constant, in which case the solution is just the Schwarzschild solution of general relativity. This result is independent of the value of the Brans-Dicke coupling constant ω .

Brans-Dicke scalar fields are perhaps of more physical interest than the conventional scalar field, since Brans-Dicke fields appear in some superstring models⁵. In Brans-Dicke theory, the scalar field is introduced by replacing the gravitational action $S_g = \frac{1}{2} \int_M R(\sqrt{-g}) d^4x$ with $\frac{1}{2} \int_M (\Phi R + \omega \Phi_{,a} \Phi^{,a} / \Phi^2) \sqrt{-g} d^4x \equiv S_g^{BD}$, where ω is the usual Brans-Dicke constant. For closed Friedman $S_g^{BD} = D \int (\Phi R + \omega \Phi^{-2} (d\Phi/dt)^2 a^3) dt$, where D is a constant and $R = 6(a''/a + (a'/a)^2 + k/a^2)$. The prime denotes time derivative. The action S_g^{BD} will be infinite unless the closed universe begins and ends in singularities, which requires $(2\omega + 3) > 0$. The finite range of t integration means S_g^{BD} will be finite or not depending on its behaviour near the singularity, where the k/a^2 term in R is negligible, so it is sufficient to look at the behaviour of S_g^{BD} for $k=0$, as $t \rightarrow 0$. In this case we have $R \propto t^A$, $\Phi \propto t^B$, where A and B are constants, and near the singularity the action goes as

$$S_g^{BD} \propto t^{3A+B-1} \quad (2)$$

The field equations for Brans-Dicke theory are given in ref. 25. In the Friedman case, the first integral of the equation containing Φ to second order in time is

$$a^3 d\Phi/dt = 1/(2\omega + 3) \int_0^t (\mu - 3p) a^3 dt + C. \quad (3)$$

When the integration constant C equals 0 we have, for $p = (\gamma - 1)\mu$, an exact power law solution with $A =$

$(2 + 2\omega(2 - \gamma))/(4 + 3\omega\gamma(2 - \gamma))$, which together imply $B + 3\gamma A = 2$, and hence equation (2) gives $S_g^{BD} \propto t^{1+3A(1-\gamma)}$. Thus the Brans-Dicke action will be finite if $C = 0$ and if $1 < \gamma < \gamma_c$, where $\gamma_c = 2 + \omega^{-1}(1 - ((2\omega + 3)/3)^{1/2})$ is obtained from $1 + 3A(1 - \gamma) = 0$. This gives the dividing line between finite and infinite actions when $\gamma \neq 1$ or $4/3$. When $\gamma = 1$, $S_g^{BD} \propto t$ and is finite; when $\gamma = 4/3$, $S_g^{BD} \propto t^{1/2}$ and is also finite; when $\gamma = 2$, $S_g^{BD} \propto t^{-1/2}$ and diverges at the singularity. Except for $\gamma = 2$, this is essentially the same behaviour we have seen in the absence of the Brans-Dicke field. This occurs because the $C = 0$ solutions are matter-dominated rather than Φ -dominated near the singularity.

In contrast, the $C \neq 0$ solutions are dominated by the Brans-Dicke scalar field near the singularity, and this leads to infinite actions. Weinberg²⁵ gives the $p=0$ solution. The asymptotic form of the general solution²⁶ for all γ when $k=0$ is $a(\tau) \propto \tau^A$, $\Phi \propto \tau^B$, where $d\tau = dt/a^{3(\gamma-1)}$ defines the 'conformal time' τ , and $A = \omega(3 + 3\omega(2 - \gamma) \pm (-\lambda))^{1/2}$, $B = (1 \pm \lambda)(1 + (2 - \gamma)\omega \pm (-\lambda))^{1/2}$, where $\lambda \equiv (1 + 2\omega/3)^{1/2}$. This is also the asymptotic form of the closed solution as $t \rightarrow 0$. Thus $S_g^{BD} \propto \tau^{B-1+3(2-\gamma)A}$ near the singularity, and a calculation shows that in all cases we have $B - 1 + 3(2 - \gamma)A = 0$, so S_g^{BD} diverges logarithmically near the singularity.

The general vacuum Brans-Dicke Friedman solutions are known in closed form, and these also have divergent S_g^{BS} for all k . When $k=0$, the solution is $a(t) = t^{1/(3(1+d))}$, $\Phi(t) = t^{d/(1+d)}$, where $d = \omega^{-1}(1 \pm (1 + 2\omega/3)^{1/2})$ and $c = d/(1+d)$. It is easily checked that $3A + B - 1 = 0$, and $S_g^{BD} \propto \ln(t) \rightarrow \infty$ near the singularity. If $k=+1$, the solution is $a^2(\tau) = (\alpha/\Phi_0) \sin 2\tau(\tan \tau)^{-\epsilon/2\alpha}$, $\Phi(\tau) = \Phi_0(\tan \tau)^{\epsilon/2\alpha}$, where $3\alpha^2 = (3 + 2\omega)\epsilon^2$, and $dt = a d\tau$. Again, this solution reduces to the $k=0$ solution near the singularity, and gives a divergent action.

Finite action in more general cosmologies

Hitherto we have considered only Friedman models, but the addition of anisotropy does not seem to change the above results significantly. For example, for perfect fluids in the Kasner-like (Bianchi Type I) models, we have $a^3 \propto t$ and $\mu \propto a^{-3\gamma} \propto t^{-\gamma}$, so $S_m \propto (\gamma - 4/3) \int \mu t dt \propto t^{2-\gamma}$ if $\gamma \neq 2$, and $\propto \ln(t)$ if $\gamma = 2$. An S^3 spatial topology Taub-like (Bianchi Type IX) exact solution for $\gamma = 2$ is given in ref. 27; a straight-forward calculation shows that S_m diverges near either singularity. On the other hand, the Kantowski-Sachs $S^2 \times S^1$ homogeneous dust-filled closed universe²⁸, with metrics $ds^2 = -dt^2 + X^2(t) dr^2 + Y^2(t) d\Omega^2$, where $X(t(\eta)) = 1 + (\eta + b) \tan \eta$, $Y(t(\eta)) = a \cos^2 \eta$, $\mu = \sec^4 \eta / a^2 [1 + (\eta + b) \tan \eta]$, $t = t_0 - a(\eta + (1/2) \sin 2\eta)$, and $a \neq 0$, $-\pi/2 < b < 0$ are constants, has finite action, since $S_m \propto \int \mu \sqrt{-g} (dt/d\eta) d\eta = \int a^2 (1 + \cos 2\eta) d\eta$, which is finite since $-\pi/2 < \eta < \pi/2$.

Some types of matter, such as electromagnetic fields, can only exist in anisotropic models. In this case we have $S_m \propto \int ((1/2) F_{ab} F^{ab} \sqrt{-g} dt d^3x \propto \int (B^2 - E^2) \sqrt{-g} dt d^3x$, where B and E are the magnetic and electric fields. For electromagnetic radiation we have $B^2 = E^2$, so $S_m = 0$. The Brill electro-vac universe²⁷ (basically a Taub universe, Bianchi Type IX, with electromagnetic fields) is a model wherein $(B^2 - E^2) \neq 0$, and a calculation shows that S_m is finite. Thus in general we would expect electromagnetic cosmological actions to be finite, and similarly for Yang-Mills actions in cosmology, since generally they correspond closely to electromagnetic actions³⁰.

As with anisotropy, inhomogeneity does not seem to change significantly the conclusions reached in the Friedman models. The dust-filled Tolman³¹ (S^3 spatial topology) models, and the Szekeres^{32,33} (S^3 and $S^2 \times S^1$ spatial topology) models both have finite action, whilst the closed $p=\mu$ and massless scalar field gravitational solitons of Belinskii³⁴ have infinite actions.

The finite-action principle is most powerful when confronted with higher-order curvature corrections to the Einstein-Hilbert lagrangian (equation (1)). Let us consider the possible quadratic

additions to equation (1) in which L_g is replaced by³⁵

$$L_Q = R + \alpha R^2 + \beta R_{ab} R^{ab} + \delta R_{abcd} R^{abcd} \quad (4)$$

where R_{ab} is the Ricci curvature, R_{abcd} the Riemann curvature, and α , β and δ are arbitrary constants. We shall neglect the problem of the new surface integral terms on ∂M arising from equation (4).

If we have an isotropic and homogeneous cosmological solution which has asymptotic behaviour $a(t) \propto t^n$ on approach to $t=0$, then the total gravitational action is

$$S_g = \frac{2n(1-2n)t^{3n-1}}{(3n-1)} + \frac{4n^2 t^{3(n-1)}}{(n-1)} \times (3\alpha + \beta + \delta + n(n-1)(12\alpha + 3\beta + 2\delta)) \quad (5)$$

The first term on the right-hand side is the contribution from the general relativity lagrangian $L_g = R$ and it diverges as $t \rightarrow 0$ for $n \leq 1/3$ as expected, (since $n = 2/3\gamma$ for Friedman universes). Now when $\alpha : \delta : \beta = 1 : 1 : -4$ the quadratic contribution to equation (4) gives rise to a complete divergence when the action is varied in four-dimensional space-time and hence the field equations (and their solutions) are identical to those obtained in general relativity. However, the presence of quadratic terms in this combination is not innocuous because the general relativity solution leads to a divergence of the quadratic contribution to L_Q as $t \rightarrow 0$ going as $S_g \propto n^3 t^{3(n-1)}$. Thus there is a divergence at $t=0$ for all $n \leq 1$, that is, for all perfect fluids with $\gamma \geq 2/3$. Since $\gamma < 2/3$ Friedman models expand forever, we see that the special combination $\alpha : \delta : \beta = 1 : 1 : -4$ gives a divergence for all perfect fluid equations of state and therefore is excluded by the finite action principle (in the Friedman case at least) even though it leaves the field equations unaffected. More generally, it is easy to check from equation (1) that quadratic terms are only admitted by the finite action principle for the special case of $n = \frac{1}{2}$ and $\beta = -2\delta/5$. Of course, there may exist other exact cosmological solutions of the theory derived from the lagrangian L_Q

which lead to closed singular cosmological models with finite action.

The very stringent constraints imposed upon gravitation theories derived from lagrangians of the form in equation (4), and those containing third and higher-order curvature invariants could be of considerable interest. Such lagrangians are predicted to arise as the leading order corrections to general relativity in the low-energy ('field theory') limit of 10- and 26-dimensional superstring theories⁵.

The fact that the universal action will be finite when the Cosmological Principle holds only if the Universe is closed in both space and time imposes strong constraints on the spatial topology. If the standard energy conditions hold and if the initial and final singularities are sufficiently strong, then a maximal hypersurface (a time of maximum expansion) must occur in any universe with a compact Cauchy surface^{2,18,19,36}. The topology of a non-flat maximal hypersurface in a closed universe satisfying the weak energy condition is strongly restricted; in fact, the only simple topologies allowed are S^3 and $S^2 \times S^1$.

Conclusions

In summary, the principle of finite action could be as useful in cosmology as it is in classical Yang-Mills theory. It provides a powerful constraint on admissible matter fields, and if the Cosmological Principle holds, it requires the Universe to be spatially closed, have special spatial topology, and be bounded in time by singularities. Moreover, constraints are imposed upon deviations from general relativity in the form of Brans-Dicke fields or additions to the Einstein-Hilbert lagrangian which are quadratic in the curvature. Conversely, a universe with nonsingular curvature invariants must be singular in its most fundamental invariant, the action.

We are grateful to M. Madsen, D. Deutsch, N. Bishop, N. P. Buchdahl, K. Kuchař and D. W. Sciama for helpful discussions. F.J.T. was supported in part by the NSF and in part by the SERC at the University of Sussex. J.D.B. acknowledges a Nuffield Foundation Fellowship.

Received 31 March; accepted 30 October 1987.

- Planck, M. *Scientific Autobiography and Other Papers* pp. 48, 180 (Greenwood, New York, 1968).
- Barrow, J. D. & Tipler, F. J. *The Anthropic Cosmological Principle* (Oxford University Press, 1986).
- Dirac, P. A. M. in *Selected Papers on Quantum Electrodynamics* (ed. Schwinger, J.) 312-320 (Dover, 1958).
- Ramond, P. *Field Theory: A Modern Primer* (Benjamin, New York, 1981).
- Green, M. B., Schwarz, J. H. & Witten, E. *Superstring Theory* Vol. 1 (Cambridge University Press, 1987).
- Jaffe, A. & Taubes, C. *Vortices and Monopoles: the Structure of Static Gauge Theories* (Birkhäuser, Boston, 1980).
- Freed, D. S. & Uhlenbeck, K. *Instantons and Four-Manifolds* (Springer, Berlin, 1984).
- Coleman, S. in *The Whys of Subnuclear Physics* (ed. Zichichi, A.) 805-941 (Plenum, New York, 1979).
- DeWitt, B. S. in *Relativity, Groups, and Topology* (eds DeWitt, C. & DeWitt, B. S.) 587-820 (Gordon & Breach, New York, 1964).
- York, J. M. *Phys. Rev. Lett.* **28**, 1082-1084 (1972).
- Gibbons, G. W. & Hawking, S. W. *Phys. Rev. D* **15**, 2752-2756 (1977).
- Hawking, S. W. in *General Relativity: An Einstein Centenary Survey* (eds Hawking, S. W. & Israel, W.) 746-789 (Cambridge University Press, 1979).
- Hawking, S. W. & Ellis, G. F. R. *The Large-scale Structure of Space-time* (Cambridge University Press, 1973).
- Narlikar, J. V. *Astrophys. J.* **5**, 67-78 (1984).
- Gott, J. R. *Nature* **295**, 304-306 (1982).
- Guth, A. *Phys. Rev. D* **23**, 347-356 (1981).
- Linde, A. D. *Phys. Lett.* **175B**, 395-400 (1986).
- Barrow, J. D., Galloway, G. J. & Tipler, F. J. *Mon. Not. R. astr. Soc.* **223**, 835-844 (1986).
- Tipler, F. J. in *Proc. 13th Texas Symposium on Relativistic Astrophysics* (ed. Ulmer, M. P.) (World Scientific, Singapore, 1987).
- Tipler, F. J. in *Proc. 26th Liège Astrophysical Colloquium* (ed. Demaret, J.) (Liège Observatory Press, Liège, 1987).
- Barrow, J. D. & Matzner, R. A. *Phys. Rev. D* **21**, 336-340 (1980).
- Tipler, F. J., Clarke, C. J. S. & Ellis, G. F. R. in *General Relativity and Gravitation* Vol. 2 (ed. Held, A.) 97-206 (Plenum, New York, 1980).
- Christodoulou, D. *Commun. math. Phys.* **105**, 337-361 (1986).
- Yodzis, P., Seifert, H. J. & Müller zum Hagen, H. *Commun. math. Phys.* **37**, 29-40 (1974).
- Weinberg, S. *Gravitation and Cosmology* (Wiley, New York, 1972).
- Gurevich, L. E., Finkelstein, A. M. & Ruban, V. A. *Astrophys. Space Sci.* **22**, 231-242 (1973).
- Barrow, J. D. *Nature* **272**, 211-215 (1978).
- Kantowski, R. & Sachs, R. K. *J. Math. Phys.* **7**, 443-446 (1966).
- Brill, D. R. *Phys. Rev.* **133**, 845-848 (1964).
- Perry, M. J. *Phys. Lett.* **71B**, 234-236 (1977).
- Bonnor, W. *Mon. Not. R. astr. Soc.* **167**, 55-65 (1974).
- Szekeres, P. *Commun. math. Phys.* **41**, 55-64 (1975).
- Bonnor, W. B. & Tomimura, N. *Mon. Not. R. astr. Soc.* **175**, 85-94 (1976).
- Belinskii, V. A. *Sov. Phys. JETP* **50**, 623-631 (1979).
- Barrow, J. D. & Ottewill, A. J. *Phys. A* **16**, 2757-2775 (1983).
- Barrow, J. J. & Tipler, F. J. *Mon. Not. R. astr. Soc.* **216**, 395-402 (1985).

Lineage-specific expression of a human β -globin gene in murine bone marrow transplant recipients reconstituted with retrovirus-transduced stem cells

Elaine A. Dzierzak, Thalia Papayannopoulou* & Richard C. Mulligan†

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

* Department of Medicine, University Hospital, University of Washington, Seattle, Washington 98195, USA

Recombinant retroviral genomes encoding a chromosomal human β -globin gene have been used to transduce murine haematopoietic stem cells in vitro. After permanent engraftment of lethally irradiated recipients with the transduced cells, the human β -globin gene is expressed at significant levels only within the erythroid lineage. These results indicate that it is possible to obtain stable expression of exogenous chromosomal DNA sequences introduced into mature haematopoietic cells in vivo via stem cell infection, and that human disorders of haemoglobin production may be more feasible candidates for somatic cell gene therapy than previously suspected.

ALTHOUGH previous studies by a number of investigators have clearly shown that retroviral vectors can be used to introduce exogenous DNA sequences efficiently into totipotent haematopoietic stem cells¹⁻³, expression of the inserted sequences has been much less well documented. Several groups have reported the expression of transferred genes in bone marrow progenitor cells before bone marrow transplantation^{4-8,11} and shortly after transplantation^{1-3,9-11}. But few studies that have focused on bone marrow transplant recipients engrafted for long periods of time have convincingly demonstrated significant levels of expression of the introduced gene^{1,2,10,11}.

To date, all of the vectors tested in bone marrow-derived cells have been designed to promote the expression of inserted complementary DNA sequences. These vectors have used either the viral transcriptional signals in the long terminal repeat (LTR) of the provirus^{1-3,5-8,10-12} or specific non-retroviral transcriptional control sequences^{4,9,11-14} to generate chimaeric RNAs. One potential explanation for the difficulties in expression encountered with the cDNA vectors is that the transcriptional sequences used lack elements that may be specifically required in haematopoietic stem cells to ensure gene expression at a subsequent stage of haematopoietic cell development. Alternatively, the chimaeric transcriptional units of the vectors may give rise to unusual messenger RNA structures which are abnormally processed or degraded. If either is the case, it might be advantageous to use sequences which contain intact cellular transcriptional units that are normally active in haematopoietic cells. Here we report the introduction of a variety of recombinant genomes encoding an intact human β -globin gene into murine haematopoietic stem cells and the reconstitution of lethally irradiated recipients with the transduced cells. We show that the human β -globin gene is expressed in transplant recipients reconstituted for long periods of time at significant levels only in the erythroid cells. We also show that the proviral transcriptional unit is active and expressed in a lineage nonspecific manner.

Infection of cells and marrow transplantation

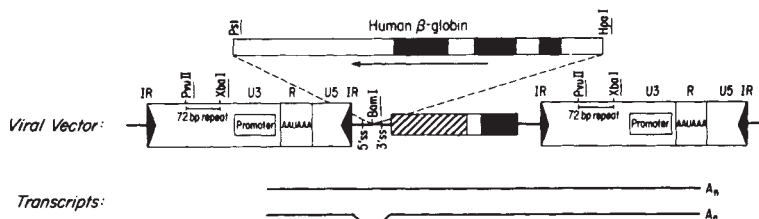
The three recombinant genomes used for this study are shown in Fig. 1. In two constructs, a 3 kilobase (kb) *HpaI*-*PstI* segment of human DNA encoding the human β -globin gene was inserted into the unique *Bam*HI site of the vectors, so that the direction of β -globin transcription is opposite to that of proviral transcription. As it was uncertain whether the presence of enhancer

element in the LTR of the vectors would affect β -globin expression *in vivo*, one construct was deleted for this sequence. The precise structure of each genome as well as the transcriptional properties of these two constructs in NIH 3T3 cells and murine erythroleukaemia cells (MEL) have been reported previously¹⁵. The source of enhancer (+) pSV(X)neo-human β -globin virus was the cell line ψ 2 3-1, which produces 1×10^5 G418-resistant 3T3 colony forming units per ml. The source of enhancer (-) pSV(X)neo human β -globin virus was ψ 2 12-1; the titre was estimated to be $>10^4$ ml⁻¹ on the basis of Southern blot analysis of infected 3T3 cells¹⁵. The third recombinant genome was constructed by the insertion of a 4.1 kb *HpaI*-*XbaI* DNA segment of the human β -globin gene into pSV(X) neo. The 1.1 kb of additional 3' flanking sequence contains a putative enhancer element¹⁶ which may be responsible for increased levels of human β -globin expression in transgenic mice (refs 17, 18 and F. Costantini, personal communication). The source of this pSV(X)neo human β -globin virus was the ψ 2-HX19 cell line which produces 2×10^5 G418 resistant colony-forming units per ml.

Bone marrow cells from C3H/HeJ female mice were prepared and infected *in vitro* as described previously³. About 5×10^4 - 4×10^6 cells were then injected into the tail vein of 104 lethally irradiated male mice. Details of the various infection and transplantation protocols used for the individual mice are shown in Table 1. One to two months after transplantation, peripheral blood samples were taken from each recipient, and DNA was prepared and analysed by the method of Southern¹⁹ for the presence of proviral sequences. From this analysis, 18 animals were found to contain cells harbouring proviral sequences (0.02-0.40 copies per cell) (data not shown), and data from eight of these mice are reported here. To determine the extent of chimaerism and the efficiency of infection, each mouse was killed 4-9 months after transplantation and DNA prepared from spleen and bone marrow was analysed by a quantitative Southern blot analysis (Fig. 1). Each DNA, when digested with *SacI* (which cuts proviral DNA once in each LTR) and probed with labelled β -globin sequences (4.1 kb *HpaI*-*XbaI* fragment of the human β -globin gene) should yield a single hybridizing fragment of 7.2 kb from the 3-1 and 12-1 proviruses or 8.3 kb from the HX19 provirus. To determine the efficiency of infection, the filter was first hybridized with the β -globin probe (Fig. 1b). It was then washed and hybridized with a Y-chromosome specific probe²⁰ to determine the extent of engraftment of female-donor bone marrow in male recipient mice (Fig. 1c). Analysis of the Y-sequence hybridizing bands in panel c indicates that each

† To whom correspondence should be addressed.

3-1 Enhancer+virus $1 \times 10^5 \text{ ml}^{-1}$



The diagram illustrates the gene transfer strategy for human β -globin. At the top, the Human β -globin gene is shown with exons represented by black boxes and introns by lines. Restriction sites PstI, HpaI, XbaI, and HhoII are indicated. Below this, the Viral Vector is shown with its components: 1R, 72 bp repeat, Promoter, AATAAA, U3, R, U5, IR, and a 5' splice site. A BamHI site is located between the U5 and IR regions. The Viral Vector is flanked by PvuII sites. The Transcripts are shown as two lines, A_n, with a V-shaped deletion between them. The Provirus is shown with the same components as the Viral Vector, but with a deletion between the U5 and IR regions, flanked by PvuII sites.

Human β -globin

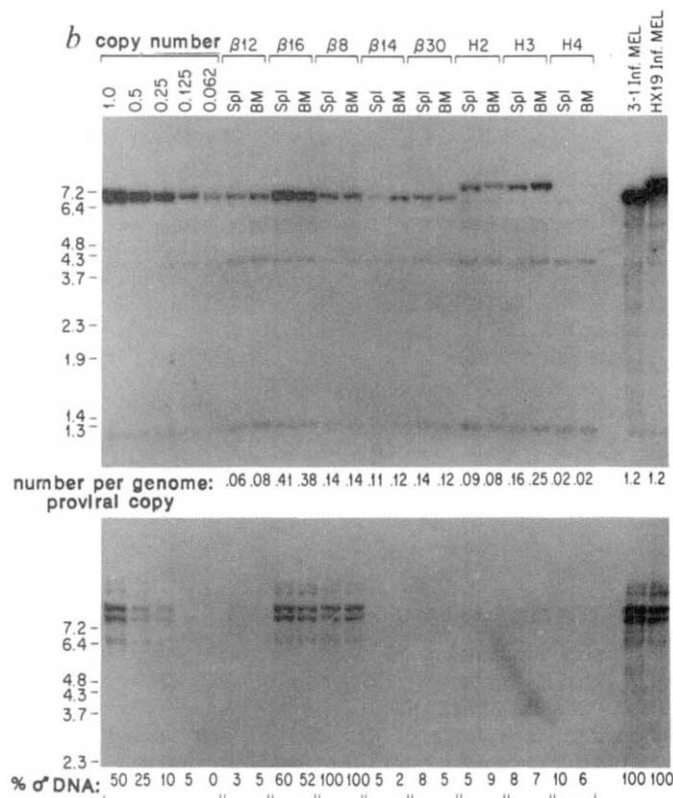
Viral Vector:

Transcripts:

A_B

A_S

DNA in 10 μ g DNA is indicated.



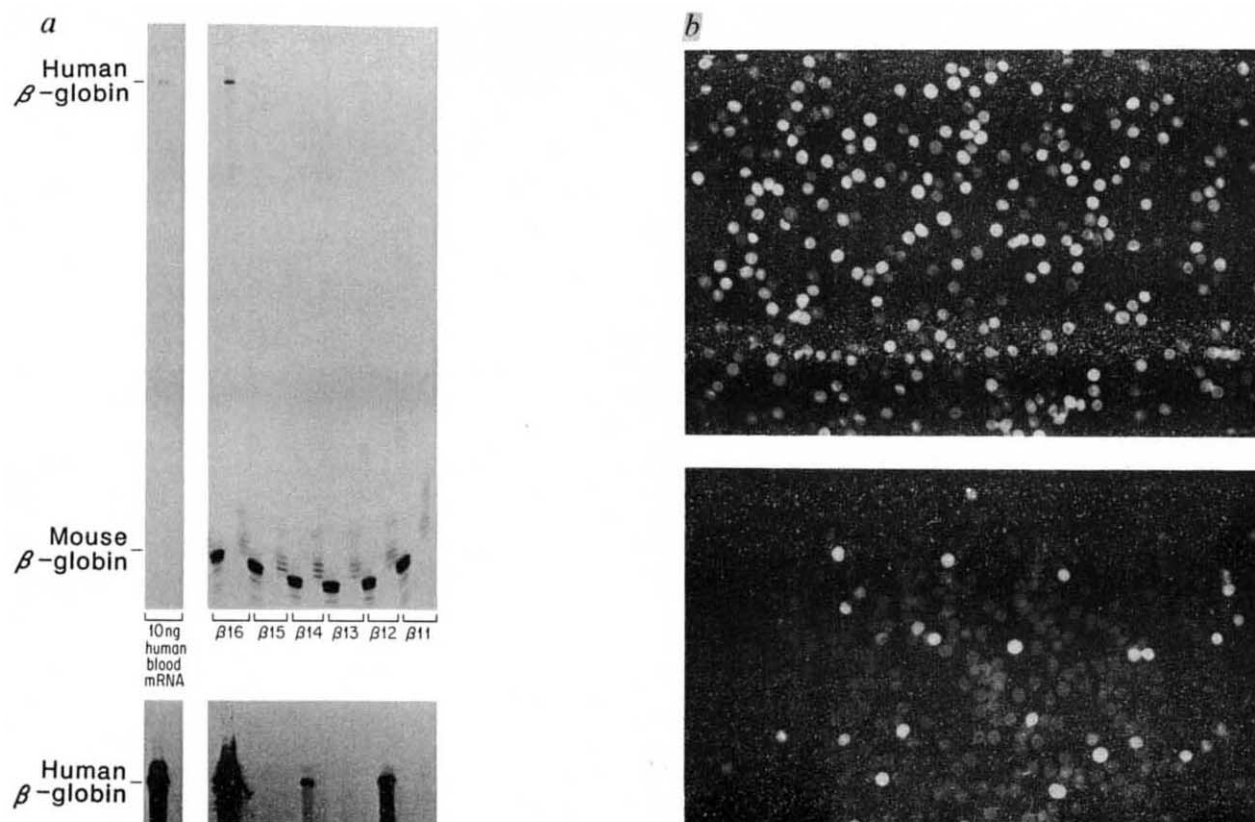


Fig. 2 Human β -globin expression in peripheral blood cells of bone marrow engrafted mice. *a*, RNA was isolated from 100 μ l of blood $\sim$ 3 months after transplantation, by the lithium-chloride method³⁴. About 1–5 μ g of total cellular RNA from each mouse was hybridized with 4×10^5 c.p.m. of 32 P-labelled human β -globin SP6 transcript and mouse β -globin SP6 transcript as previously described¹⁵ except that hybridizations were carried out at 55 $^{\circ}$ C. Human blood total cellular RNA (10 ng) and 10 ng of MP10 total cellular RNAs³⁵ were used as controls for the human β -globin and neo probes respectively. After RNase digestion, protected fragments were denatured and separated on a 6% acrylamide denaturing gel. The sizes of protected fragments are as follows: 351 nucleotides for human β -globin, 48 for mouse β -globin and 168 for neo. The top panel shows an exposure time of 20 min, and the lower panel shows an exposure of 2 days. *b*, A monoclonal antibody against the human β -chain was used to stain blood cells as previously described²⁴. Preparations of blood cells from mice B16 (top panel) and B12 (bottom panel) are shown. Anti-human β -chain antibody-stained cells from other mice are described in Table 2. All negative samples showed no fluorescent cells. The threshold sensitivity of the antibody has been shown to be less than 1 pg of globin protein per cell³⁶.

recipient was engrafted with donor-derived cells, although the extent of donor engraftment varied from 40% in mouse B16 to 98% in mouse B14 (the extent of engraftment of mouse B8 could not be determined, as it was engrafted with infected male C3H/HeJ bone marrow). Analysis of the gels shown in panel *b* indicates that each mouse was engrafted with retrovirally transduced cells, in that each spleen and bone marrow DNA sample contained a 7.2 kb or 8.3 kb β -globin hybridizing fragment that co-migrated with the band from the authentic retroviral construct. Hybridization of the same blots with a labelled neo probe confirmed these results and additional digestions of the DNAs from cells infected with the enhancer (–) construct indicated that the expected proviral structure was transmitted (data not shown). As the human β -globin probe cross-hybridizes to the mouse β -globin gene, densitometric scanning of the mouse-derived sequences was used to normalize the amount of DNA loaded per lane. The copy number of the integrated proviral sequences was then determined by comparison of the band intensities to those representing known quantities of plasmid DNA. Based on this analysis, the efficiency of infection ranged from 2% (in mouse H4, where >90% of cells were donor derived) to 100% (in mouse B16, where only 40% of cells were donor derived).

β -Globin gene expression in recipients

To determine whether the β -globin-containing proviruses were transcriptionally active, RNA was first prepared from the peripheral blood of three of the recipients analysed in Fig. 1

(B12, B14 and B16) $\sim$ 3 months after transplantation and analysed by an SP6 protection assay²¹ (Fig. 2*a*). Blood RNAs from three additional recipients subsequently shown to contain no (B13 and B15) or very few (B11) detectable transduced cells were also analysed. To characterize human β -globin RNA transcripts, an SP6 probe was used that extended from 128 base pairs upstream from the cap site to the *Bam*HI site in the second exon^{15,22}. Hybridization to this probe assays for the correct 5' end of human β -globin RNA. To compare the level of human β -globin mRNA to the level of endogenous mouse β -globin mRNA, we used a mouse-specific probe that protects 48 nucleotides in the first exon when hybridized with authentic mouse β -globin mRNA^{15,23}. As shown in the SP6 protection analysis in Fig. 3*a*, the three proviral positive recipients show significant but varying levels of human β -globin expression (see lower panel for longer exposure). Mouse B16 shows the highest level of expression as compared to the signal in 10 ng of human blood RNA. B14 and B12 express lower but detectable levels. The corresponding level of mouse globin mRNA is also shown for each recipient.

At this time erythrocytes were also isolated from each recipient and fluorescein stained with anti-human β -chain monoclonal antibody²⁴ to determine the percentage of cells which produce human β -globin protein. In each sample from the RNA-positive animals, brightly fluorescent cells could be detected. Figure 2*b* shows that 50% of erythrocytes from B16 (top) and 10% of erythrocytes from B12 (bottom) stain brightly. Based on this estimation of copy number from the staining assay and Cerenkov

Table 1 Bone marrow engrafted mice

Animal	B12	B16	B8	B14	B30	H2	H3	H4
<i>In vivo</i> 5FU-treatment of bone marrow donor (days)	2	2	2	2	6	6	6	6
Virus	3-1	3-1	3-1	12-1	12-1	HX19	HX19	HX19
<i>In vitro</i> G418 selection (1 mg ml ⁻¹)	—	+	+	—	—	—	—	—
Number of bone marrow cells transplanted	2.5 × 10 ⁵	3.5 × 10 ⁵	4 × 10 ⁶	2.5 × 10 ⁵	1 × 10 ⁵	5 × 10 ⁴	5 × 10 ⁴	5 × 10 ⁴
Time post-transplant of analysis (months)	4.5	4	9	4.5	5.5	4.5	4.5	4.5

Infection and transplantation protocol for bone marrow engrafted mice. C3H/HeJ female mice (7–12 weeks, Jackson) were injected with 5-fluorouracil (Adria) at 140–150 mg per kg body weight. Bone marrow was collected either 2 or 6 days after injection as described previously³. Between 10⁵ and 10⁶ bone marrow cells were co-cultured with a monolayer of ψ 2 3-1, ψ 2 12-1 or ψ 2 HX19 cells for 48 h as described previously³. Recovery of viable cells ranged from 10% to 50% for unselected bone marrow and 5% to 18% for bone marrow kept in culture an additional 48 h in G418 (1 mg ml⁻¹). Cell numbers used for injection into lethally irradiated (1,100 rad) C3H/HeJ male mice (aged 7–20 weeks) are as indicated above. Handling of cells for injection, irradiation and housing and handling of animals was as previously described³. Animals were killed at the indicated times post-transplantation. When serum from engrafted animals was tested for the presence of helper virus by XC assay³² two out of ten animals showed 1–2 XC plaques, but only after multiple passages of 3T3 cells exposed to the serum. No G418 resistant colonies were found with any of the sera tested after multiple passages indicating no spread of the recombinant human β -globin viruses.

counting of protected fragments isolated from the SP6 gel represented in Fig. 2a, the level of human β -globin mRNA in those erythrocytes positive for human β -globin protein 3 months after transplantation is ~4.8% of the level of mouse β -globin mRNA in B16 and 1.3% of the level of mouse β -globin mRNA in B12 (see legend to Table 2).

When peripheral blood RNA was prepared from the eight animals in Fig. 1 at the time of killing (4–9 months after transplantation), human β -globin expression was again analysed by SP6 protection for quantitation of RNA and by immunofluorescence for detection of protein. The percentage of erythrocytes in each blood sample stained with the human β -chain specific antibody (Table 2) approximated the percentage of spleen and bone marrow cells bearing proviruses as determined by the Southern blot analysis (Fig. 1) indicating that the copy number in the erythroid lineage is equivalent to that in other haematopoietic lineages. Thus, erythroid cells derived from bone marrow cells bearing a proviral integrant express human β -globin protein, with the level of human β -globin mRNA expression varying between 0.4% and 4.0% of endogenous mouse β -globin mRNA levels when considered on a single-copy-per-genome basis (Table 2). The inclusion of the extra human β -globin DNA sequences in HX19 did not appear to dramatically increase the amount of human β -globin transcription.

Tissue specificity of expression

To determine whether human β -globin expression was restricted to the erythroid compartment of the engrafted mice, haematopoietic cells from the eight transplant recipients were fractionated into cell populations representing different lineages as previously described³. For this analysis, the cell populations included: (1) peripheral blood; (2) total spleen; (3) total bone marrow; (4) spleen T cells; (5) spleen B cells; (6) lymph node T cells; (7) lymph node B cells; and (8) primary culture of macrophages. DNA and RNA were prepared from each fraction. To assess whether all haematopoietic lineages in the engrafted mice were derived from the same infected totipotent stem cell(s), we performed Southern blot analysis of DNA from each of these cell populations using restriction endonucleases *Dra*I or *Bam*HI which produce DNA fragments diagnostic for the unique site of provirus integration in genomic DNA. When probed for human β -globin or neo hybridizing sequences, each recipient showed a single (or very few) identical proviral integrant(s) in each cell population with about the same percentage contribution of infected cells in each lineage (data not shown). Thus, we concluded that a single (or very few) infected stem cell(s)

contributed to all the haematopoietic populations in the long-term engrafted mice and that our RNA analysis would correctly reflect the activity of the transduced β -globin gene in mature haematopoietic cells derived from the infected totipotent stem cell(s). RNA prepared from each cell fraction was then examined by SP6 protection analysis (Fig. 3). Because, in some cases, the low amount of RNA obtained from a cell population precluded obtaining an accurate determination of the concentration of RNA in the sample, a preliminary SP6 analysis using a human γ -actin probe²⁵ which protects a 68-nucleotide fragment of mouse γ -actin, was used to normalize the amount of RNA. The mouse β -globin probe was used to verify the cell populations (i–viii) as erythroid or non-erythroid and the human β -globin and neo probes to detect expression of the transduced genes. Fractionation analysis of mouse B16, shown in panel a, indicates that human β -globin mRNA expression was indeed restricted to cell populations representing the erythroid lineage, as human β -globin expression parallels the levels of mouse β -globin mRNA expressed in blood, spleen and bone marrow. Only very small amounts of human β -globin mRNA were found in B and T lymphocytes or in macrophages. Although the level of human β -globin mRNA expression in animal B12 (panel b) was much lower, again the expression was restricted to erythroid cells. The pattern of RNA expression in animals H2, H3 and H4 was also identical to this pattern (data not shown). Because mouse B14 was engrafted with cells infected with enhancer (–) proviral

Fig. 3 (Opposite) Lineage analysis of human β -globin RNA expression. a, Mouse B16; b, mouse B12; c, mouse B14; and d, mouse B30. RNAs from various haematopoietic tissues and cell populations enriched for lymphocytes or macrophages were prepared by the lithium-chloride method³⁴. Due to low RNA yields from some enriched populations, RNA was quantitated by hybridization with ³²P-labelled SP6 human γ -actin transcript. This probe protects 68 nucleotides of mouse γ -actin. About 0.25–3 μ g of each RNA was hybridized to an excess of human β -globin and neo SP6 RNA and ~25–300 ng of each RNA was hybridized to an excess of mouse β -globin and actin SP6 RNA. Protection analysis is as described in Fig. 2. The radioactive bands migrating lower than the expected neo-protected fragment in the spleen of mouse B14 were not observed in subsequent analyses. The larger neo fragment in some of the longer exposures was undigested probe. Exposure times for the top panels (human β -globin and neo probes) of mice B16, B14 and B30 were 14, 14 and 41 h whereas the lower panel exposure times were 12, 14 and 21 days respectively. Exposure times for mouse β -globin and actin probes for B16, B14 and B30 were 14 h, 14 h and 2 days respectively. Exposure times for B12 were 12 days for human β -globin and neo probes and 8 h for mouse β -globin and actin probes.

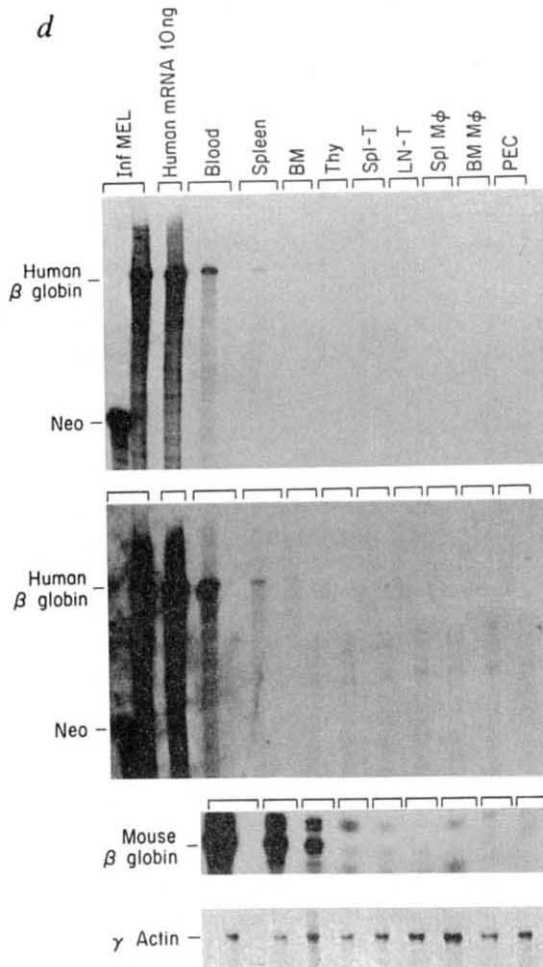
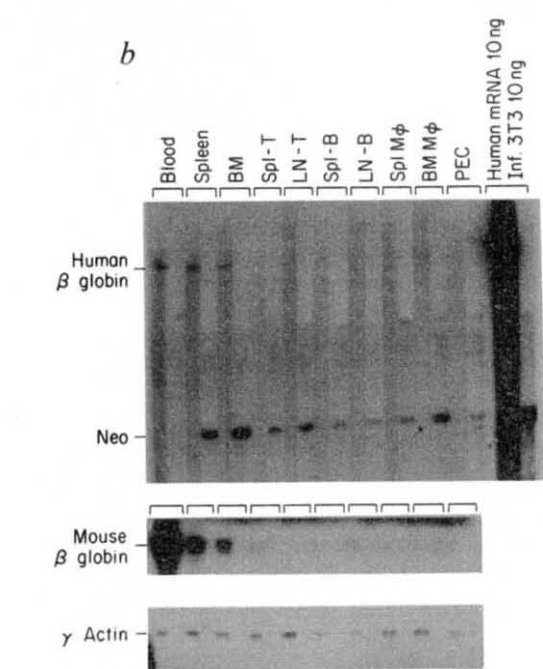
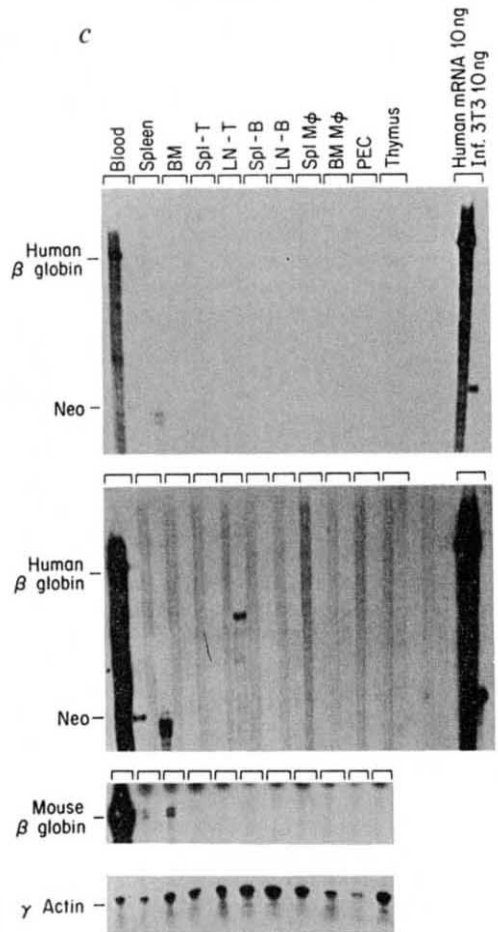
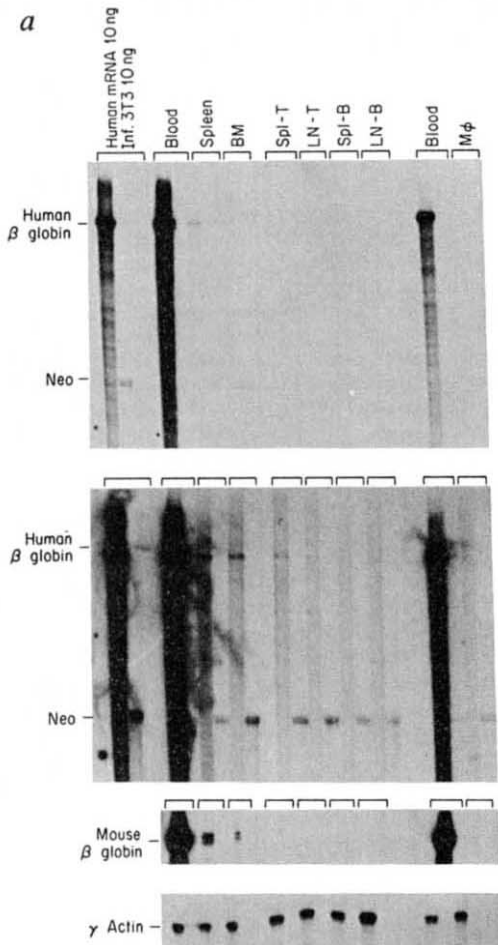


Table 2 Human β -globin expression in bone marrow engrafted mice

	B12	B16	B8	B14	B30	H2	H3	H4
H β pg μg^{-1}	0.43	5.04	2.64	0.86	0.43	3.61	2.02	1.20
M β pg μg^{-1}	1,160	905	754	754	870	1,310	1,430	1,508
% H β expression	0.04	0.56	0.35	0.10	0.05	0.28	0.14	0.08
% Fluorescent cells in peripheral blood	10	50	ND	<10	<10	2-3	13	2-4
Proviral copy number per cell	0.07	0.40	0.14	0.11	0.13	0.09	0.20	0.02
% H β expression at one copy per cell	0.6	1.4	2.5	0.9	0.4	3.3	0.7	4.0

The steady state level of human β -globin (H β) mRNA and mouse β -globin (M β) mRNA in the blood cells of engrafted mice at the time of killing as determined by quantitative SP6 protection analysis. Protection analysis for RNA samples was as described in the legend to Fig. 3b except that 5×10^6 c.p.m. of labelled probes were added to 200–900 ng of each blood RNA sample to ensure probe excess. Amount of β -globin transcript was obtained by Cerenkov counting of protected labelled probe fragments cut from SP6 gels. The c.p.m. data for each fragment were converted to pg of transcript taking into account specific activity (810 Ci mmol^{-1}), the GTP content of the protected fragment and the size of the complete β -globin transcript (600 nucleotides). This value was then divided by the quantity of total cellular RNA (μg) used for analysis. The level of endogenous mouse β -globin RNA in these animals is consistent with the published values^{17,29} when the per cent reticulocyte RNA contributing to total blood cell RNA is considered, except in the case of the analyses on mice B16 and B12 at 3 months post-transplant, where the quantities of endogenous mouse β -globin RNA measured were low (most likely due to an insufficient quantity of mouse β -globin probe). This may result in an overestimate of human β -globin expression in these mice at 3 months post-transplant. Reticulocyte counts on blood of engrafted mice yielded an average of 3–5%. As we have determined that an equivalent efficiency of hybridization is achieved with both probes during annealing, the per cent H β expression was calculated by $(\text{H}\beta / \text{M}\beta) \times 10^2$. Per cent fluorescent cells was determined as described in text (ND, not determined) and the proviral copy number was taken from Fig. 2. The per cent H β expression at one copy of provirus per cell was calculated by per cent H β expression $\div$ proviral copy number.

construct, it was of interest to determine whether the lack of viral transcriptional enhancer sequences affected the tissue specificity of human β -globin expression. As can be seen in panel c, the degree of tissue specificity of human β -globin expression is identical to, or perhaps more restricted than that in the mice engrafted with cells infected with the enhancer (+) constructs. Analysis of the other enhancer (–) mouse, B30 (panel d), confirmed the observation that the degree of tissue specificity of β -globin expression appears to be even greater than with the enhancer (+) virus.

Provirus transcription

Although previous studies had suggested that the Mo-MLV proviral transcriptional unit is not highly or reproducibly active in haematopoietic cells derived from transduced stem cells^{1,2,10}, we also examined the level of neo RNA in each cell population to determine whether the inserted β -globin sequences might affect proviral transcription. In contrast to results with other constructs, neo RNA could be detected (Fig. 3). In the case of mouse B12 and B16, the level of neo RNA was ~1–2% of the level of neo RNA found in NIH 3T3 cells containing a single proviral copy of the MP10 (ref. 26) or 3-1 genome. But in contrast to the tissue specificity of the human β -globin sequences, the proviral RNA was found in the lymphocyte and macrophage populations (Fig. 3a and b). Northern blot analysis demonstrated both unspliced and spliced transcripts (transcript ratio of ~1:4) when probed with labelled neo sequence and confirmed the levels of expression in the different lineages (data not shown). In the case of mouse B14 infected with the enhancer (–) virus, neo RNA was also detected, but only in erythroid cells (Fig. 3c). Again, the level of expression was ~1% of the level of expression found in infected NIH 3T3 cells. In the case of the enhancer (–) constructs, the tissue specificity of neo transcription may be integration-site dependent, as no neo expression was observed in mouse B30 (Fig. 3d).

Conclusions and prospects

The results presented here indicate that it is possible to obtain the stable and qualitatively appropriate expression of a normal cellular gene after its introduction into haematopoietic cells *in vivo*, via retrovirus-mediated transduction of stem cells and bone marrow transplantation. In contrast to previous studies, which have used viral constructs in which cDNA sequences are placed under the control of viral transcriptional signals^{1,2,10,11}, the recombinant retroviral genomes used here encoded an intact,

chromosomal human β -globin gene. The gene was found to be expressed exclusively in the erythroid cells of all transplant recipients, long after bone marrow transplantation, whether or not the viral transcriptional control signals in the vector were present. In keeping with our previous studies of human β -gene expression in MEL cells¹⁵, the level of expression varied less than 10-fold among different proviral integrants.

Although expression of the transferred human β -globin gene was strikingly tissue specific, the absolute level of β -globin transcription was low. Several recent studies suggest, however, that the low level of expression observed reflects the specific segment of human DNA used, rather than any unique problem of expression in the stem-cell gene transfer system. First, similar levels of expression (when considered on a single-copy-per-genome basis) have been obtained by a number of investigators after transfer of the same or similar segments of the human β -globin gene into MEL cells^{15,27,28} or into the mouse germ line^{17,29}. More importantly, Grosfeld and co-workers³⁰ have recently identified two additional sequences, located 50 kb upstream and 20 kb downstream of the β -globin structural gene which, when juxtaposed to a minimal β -globin containing DNA segment, dramatically boost expression to qualitatively normal levels in transgenic mice containing a single copy of the sequences, independent of the chromosomal site of insertion. These findings suggest that the segments of DNA used in our study still lack some sequence elements which play an important role in the proper expression of the β -globin gene and that it will be possible to significantly increase expression over the levels we have reported here.

In the light of previous studies from a number of laboratories that have indicated that proviral transcriptional signals are most often not used in cells derived from transduced haematopoietic stem cells, the finding that the proviral genome was also expressed in all mice examined suggests that the inserted human β -globin sequences may have some effect on viral transcription. This is almost certainly true in the case of the enhancer (–) constructs, as viral transcription was restricted to erythroid cells. In the case of the enhancer (+) constructs, the mechanism by which the inserted β -globin sequences could activate proviral transcription is less clear, as proviral transcription was detected in all cell types tested.

As the level of proviral transcription was low, and parallel experiments were not performed with vectors containing no inserts, no firm conclusions regarding the effects of the β -globin gene on viral transcription can be made. Nevertheless, it is

intriguing to speculate that the effect of β -globin sequences on proviral transcription may reflect a normal mechanism that exists in stem cells to 'poise' certain genes (for example, the globin genes) for eventual expression at a later developmental stage. This mechanism could involve the act of transcription, or simply the generation of a specific chromatin configuration through the binding of specific transcriptional factors. Genes lacking the necessary *cis*-acting DNA sequences to become 'poised', may then not be able to be expressed at later stages of haematopoietic cell differentiation. Experiments are now in progress to determine whether *cis*-acting sequences within the β -globin gene are indeed responsible for the activation of proviral transcription, to understand the mechanism of activation, and to establish whether other cellular genes employ a similar mechanism to regulate their own expression in haematopoietic cells.

The data also has a number of important general implications for both basic studies of gene expression in haematopoietic cells and for potential clinical applications of gene transfer. Previous studies aimed at expressing genes within the haematopoietic system via stem-cell infection have all relied on the use of cDNA expression vectors^{1,2,10,11}. The current studies suggest an alternative approach to gene expression that relies on the use of transcriptional signals normally used in haematopoietic cell differentiation. If specialized signals are indeed necessary in stem cells to provide for the subsequent tissue specific expression

of particular genes, the relevant portions of the gene could be incorporated into vectors such as those described here, either to directly promote the expression of inserted sequences, or to activate the proviral transcriptional unit itself.

Finally, it is interesting to note that historically, disorders of haemoglobin production such as sickle cell anaemia and β thalassaemia were considered to be the best candidates for gene therapy, because of the wealth of information available about normal and abnormal haemoglobin gene expression³¹. But as investigators began to experience difficulty in obtaining the proper qualitative and quantitative expression of cloned globin genes *in vitro* and *in vivo*^{15,17,27-29}, attention shifted to the potential treatment of other diseases, most notably the severe combined immunodeficiencies (SCID), which would in theory require a less stringent recapitulation of normal gene expression to achieve clinical benefit. In conjunction with the recent studies reported by Grosveld and co-workers³⁰, the data presented here suggests that it may now be warranted to reconsider disorders of haemoglobin production as potential candidates for gene therapy as well.

We thank Phil Sharp, David Baltimore, Stuart Orkin, David Housman and David Nathan for critically reviewing the manuscript, Ihor Lemischka and Mariann Grossman for help in various stages of the work, and the Maniatis laboratory for many kind gifts of plasmid DNAs. This work was supported by the NIH. E.A.D. is a fellow of the Leukemia Society.

Received 19 October; accepted 27 November 1987.

1. Dick, J. E., Magli, M. C., Huszar, D., Phillips, R. A. & Bernstein, A. *Cell* **42**, 71-79 (1985).
2. Keller, G., Paige, C., Gilboa, E. & Wagner, E. F. *Nature* **318**, 149-154 (1985).
3. Lemischka, I. R., Raulet, D. H. & Mulligan, R. C. *Cell* **45**, 917-927 (1986).
4. Joyner, A., Keller, G., Phillips, R. A. & Bernstein, A. *Nature* **305**, 556-558 (1983).
5. Gruber, H. E. *et al. Science* **230**, 1057-1061 (1985).
6. Belmont, J. W. *et al. Nature* **332**, 385-387 (1986).
7. Chang, S. M. W. *et al. Molec. cell. Biol.* **7**, 854-863 (1987).
8. Miller, A. D. & Hock, R. A. *Nature* **320**, 275-277 (1986).
9. Lim, B., Williams, D. A. & Orkin, S. H. *Molec. cell. Biol.* (in the press).
10. Eglitis, M. A., Kantoff, P., Gilboa, E. & Anderson, W. F. *Science* **230**, 1395-1398 (1985).
11. Bowtell, D. D. L., Johnson, G. R., Kelso, A. & Cory, S. *Molec. Biol. Med.* (in the press).
12. Williams, D. A., Orkin, S. H. & Mulligan, R. C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2566-2570 (1986).
13. Kantoff, P. W. *et al. J. exp. Med.* **166**, 219-234 (1987).
14. McIvor, R. S. *et al. Molec. cell. Biol.* **7**, 838-846 (1987).
15. Cone, R. D., Weber-Benarous, A., Baorto, D. & Mulligan, R. C. *Molec. cell. Biol.* **7**, 887-897 (1987).
16. Kollias, G., Hurst, J., deBoer, E. & Grosveld, F. *Nucleic Acids Res.* **15**, 5739-5747 (1987).
17. Townes, T. M., Lingrel, J. B., Chen, H. Y., Brinster, R. L. & Palmiter, R. D. *EMBO J.* **4**, 1715-1723 (1985).

18. Kollias, G., Wrighton, N., Hurst, J. & Grosveld, F. *Cell* **46**, 89-94 (1986).
19. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
20. Lamar, E. E. & Palmer, E. *Cell* **37**, 171-177 (1984).
21. Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035-7056 (1985).
22. Charnay, P. *et al. Cell* **38**, 251-263 (1984).
23. Charnay, P., Mellon, P. & Maniatis, T. *Molec. cell. Biol.* **5**, 1498-1511 (1985).
24. Stamatoyannopoulos, G. *et al. Blood* **61**, 530-539 (1983).
25. Enoch, T., Zinn, K. & Maniatis, T. *Molec. cell. Biol.* **6**, 801-810 (1986).
26. Barklis, E., Mulligan, R. C. & Jaenisch, R. *Cell* **47**, 391-399 (1986).
27. Chao, M. V., Mellon, P., Charnay, P., Maniatis, T. & Axel, R. *Cell* **32**, 483-493 (1983).
28. Wright, S., deBoer, E., Grosveld, F. G. & Flavell, R. A. *Nature* **305**, 333-336 (1983).
29. Chada, K. *et al. Nature* **314**, 377-380 (1985).
30. Grosveld, F., van Assendelft, G., Greaves, D. & Kollias, G. *Cell* (in the press).
31. Bunn, H. F. & Forget, B. G. *Hemoglobin: Molecular, Genetic and Clinical Aspects* (Saunders, Philadelphia, 1986).
32. Rowe, W. P., Pugh, W. E. & Hartley, J. W. *Virology* **42**, 1136-1139 (1970).
33. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
34. Auffray, C. & Rougeon, F. *Eur. J. Biochem.* **107**, 303-314 (1980).
35. Soriano, P., Cone, R. D., Mulligan, R. C. & Jaenisch, R. *Science* **234**, 1409-1413 (1986).
36. Bigbee, W. L., Branscomb, E. W., Weintraub, H. B., Papayannopoulou, Th. & Stamatoyannopoulos, G. *J. Immun. Meth.* **45**, 117-127 (1981).

A tissue-specific endothelial cell molecule involved in lymphocyte homing

Philip R. Streeter, Ellen Lakey Berg, Bich Tien N. Rouse, Robert F. Bargatzte & Eugene C. Butcher

Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA and Veterans Administration Medical Center, Palo Alto, California 94304, USA

An endothelial cell surface molecule that is selectively expressed in mucosal organs is required for lymphocyte homing to mucosal lymphoid tissues. This 'vascular addressin' appears to function as a tissue-specific marker or address signal for recognition by lymphocytes circulating in the blood.

ENDOTHELIAL cell differentiation has been proposed as a factor regulating the selective extravasation of blood-borne cells in a variety of systems, for example in the metastasis of non-lymphoid and lymphoid tumours¹⁻³, the migration of monocytes and neutrophils to sites of tissue inflammation^{4,5}, and the trafficking of lymphocytes from one organ to another via the bloodstream^{6,7}. In the latter case, vascular differentiation is

clearly important. Lymphocytes leave the blood by adhering to and migrating through the endothelium lining specialized venules, particularly although not exclusively the morphologically distinct high endothelial venules (HEVs) found in organized lymphoid tissues and sites of chronic inflammation^{6,7}. Also, the extravasation of distinct lymphocyte subsets can exhibit remarkable tissue-specificity (see for example refs 6-14;

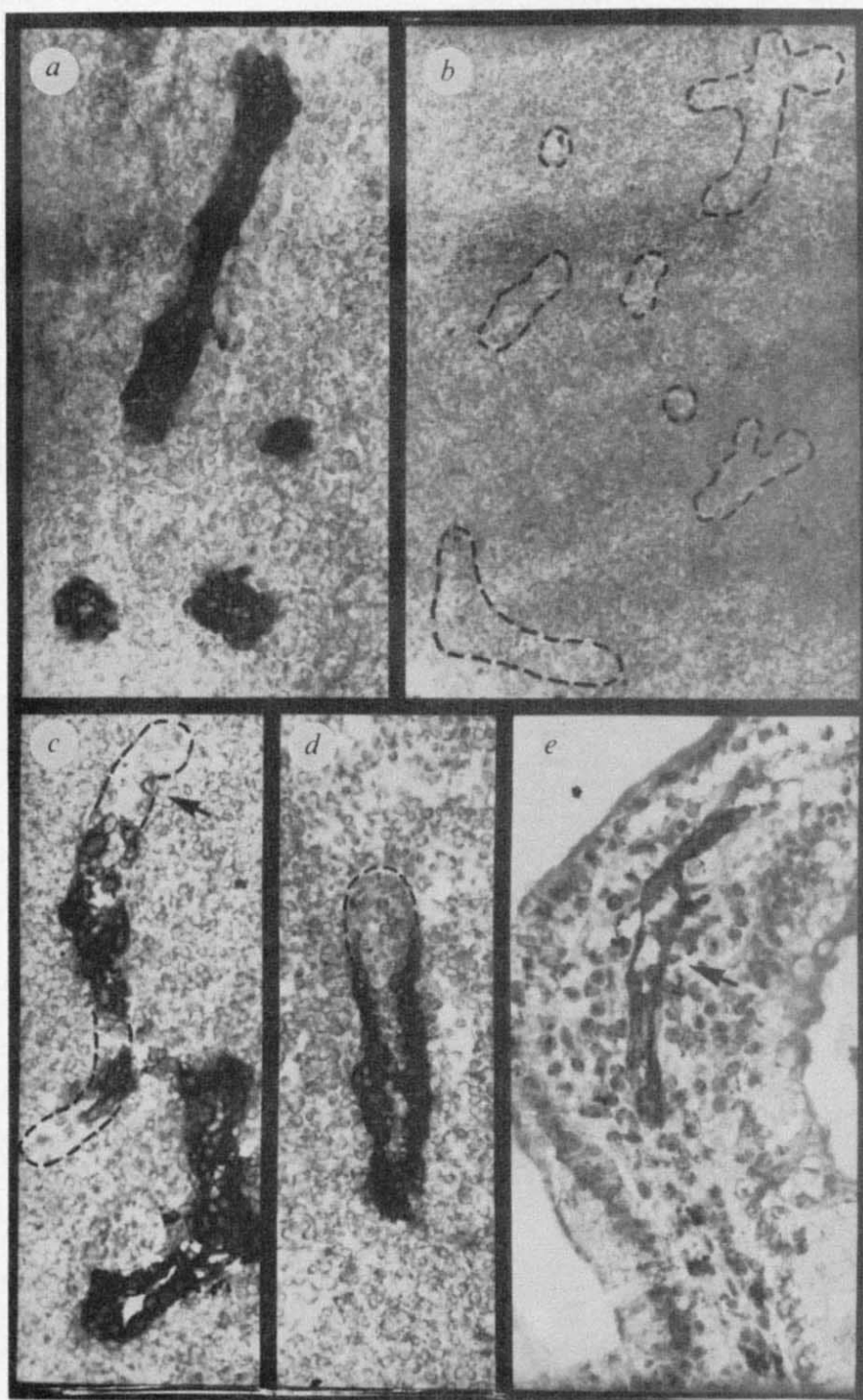


Fig. 1 Tissue-specific expression of the MECA-367 antigen. *a*, In immunoperoxidase analyses, MECA-367 intensely stains all HEVs in the organized mucosal lymphoid organs, the Peyer's patches ($\times 400$). *b*, In contrast, the antibody fails to stain most HEVs in peripheral (axillary, brachial, popliteal and inguinal) lymph nodes ($\times 200$; negative vessels are outlined). *c* and *d*, In mesenteric lymph nodes ($\times 400$), most HEVs or segments of HEVs stain intensely and uniformly with MECA-367; some contain a mixture of positive and negative HECs, giving a mottled appearance; and others are negative (negative HEV segments are outlined). The arrow illustrates staining of a single HEC in a high endothelial venule segment that is otherwise negative for MECA-367. *e*, Staining of a section through a villus of the small intestine reveals a MECA-367⁺ venule in the extralymphoid lamina propria (arrow; $\times 500$). **Methods.** mAb production: brachial, axillary, inguinal, and mesenteric lymph nodes from 8–12 week old BALB/c mice were pooled in Hank's balanced salt solution (HBSS), minced, gently pressed between glass microscope slides to release lymphocytes, and passed through nitex nylon mesh (Sullivan's, San Francisco). The stromal elements which remained on the mesh were collected, treated for 10 min (37°C) with HBSS containing 0.32 mg of collagenase per ml (5 ml per mouse), washed, again passed through nylon mesh, and used for immunization of Wistar rats with alum as adjuvant. Immunization and hybridoma production involved standard techniques as described^{21–23}, and the fusion partner was the mouse myeloma Sp2/O (American Type Culture Collection, Rockville, Maryland). Both MECA-367 and MECA-89 are rat IgG2a's. Immunoperoxidase staining: a two-stage immunolabelling technique was performed as described³⁷, and sections were lightly counterstained with haematoxylin. Controls included staining with an irrelevant class matched (rat IgG2a) antibody, Hermes-1 (ref. 21). Cross-blocking analysis of MECA-89 and MECA-367: acetone-fixed frozen sections of Peyer's patches were preincubated with either MECA-89, MECA-367, control antibody Hermes-1, or medium; washed in phosphate buffered saline (PBS), and stained with biotinylated-MECA-89, -MECA-367, or -Hermes-1, followed by texas red-avidin (Cappel, diluted 1:75 in PBS). All antibodies were used at $100\text{ }\mu\text{g ml}^{-1}$. Preincubation with MECA-89 completely blocked staining with MECA-89, without influencing staining with MECA-367; and similarly, MECA-367 blocked itself but not MECA-89. Control results were as expected. HEC isolation: viable HECs and HEV fragments were released from partially purified mesenteric lymph node stromal elements (see above) by treatment for 30 min (37°C) with HBSS containing 0.32 mg of collagenase and 0.4 mg of dispase per ml (5 ml per mouse). HECs and HEV fragments stained intensely with MECA-89 and MECA-367, but not with class-matched control antibodies.

reviewed in ref. 15), determined at least in part by selective endothelial cell recognition^{16–18}. Functional studies have revealed the existence of at least three independent lymphocyte-HEV recognition systems, mediating lymphocyte binding to HEVs in peripheral lymph nodes, in mucosal lymphoid tissues (intestinal Peyer's patches and appendix) and in the synovium of inflamed joints^{16–18}. Lymphocyte surface 'homing receptors' involved in these tissue-specific interactions with endothelial cells have been identified^{17,19–22}.

In an effort to identify endothelial cell recognition elements involved in tissue-specific lymphocyte-HEV interactions, we

have produced monoclonal antibodies (mAbs) against mouse HEVs. Two antibodies, described here, define an endothelial cell surface antigen present on HEVs and other vessels that support lymphocyte extravasation in mucosal tissues, but absent from such vessels in non-mucosal sites. We provide evidence that this antigen functions as a mucosa-specific endothelial cell position marker and recognition element for circulating lymphocytes, and that it is required for the selective extravasation of lymphocytes in mucosal lymphoid tissues. Because its apparent role is to convey positional (tissue address) information to circulating lymphocytes, we propose to call this mucosa-

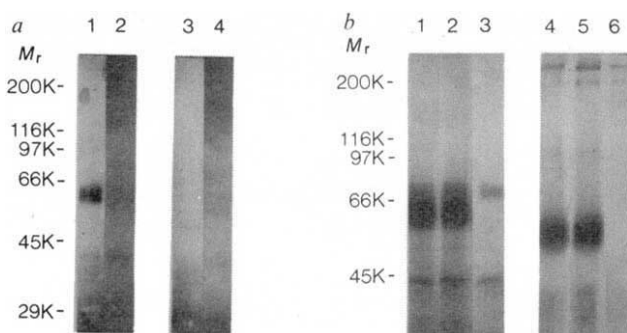


Fig. 2 MECA-367 binds a 58–66K protein that is also recognized by MECA-89. *a*, Western blot analyses of the affinity-isolated MECA-367 antigen probed with MECA-367 (lane 1) or the isotype-matched negative control antibody (Hermes-1; lane 2) illustrating the specific antigen of approximate M_r 62K. The band detected by MECA-367 is not present in material eluted from a control antibody affinity column (lanes 3 and 4, probed with MECA-367 and Hermes-1, respectively). *b*, SDS-PAGE analysis of the iodinated MECA-367 antigen. Iodinated antigen was immunoprecipitated with MECA-367 (lanes 1 and 4), MECA-89 (lanes 2 and 5), or the negative control antibody Hermes-1 (lanes 3 and 6), and electrophoresed under reducing (1–3) or non-reducing (4–6) conditions in a 10% SDS-PAGE gel.

Methods. MECA-367 antigen isolation. The MECA-367 antigen was isolated from a lysate of mesenteric lymph node stroma by immunoaffinity chromatography. Briefly, lysates were prepared by incubating BALB/c mesenteric lymph node stroma (no collagenase treatment) in lysis buffer (2% Nonidet P-40, 150 mM NaCl, 1 mM $MgCl_2$, 0.02% NaN_3 , 10 $\mu g\ ml^{-1}$ Aprotinin, 10 $\mu g\ ml^{-1}$ Pepstatin, 10 $\mu g\ ml^{-1}$ Leupeptin, 1 mM phenyl methyl sulphonyl fluoride (PMSF) in 50 mM Tris-HCl pH 7.4) for 90 min on ice. Lysates were clarified by centrifugation at 4°C, passed through 0.2 μm nitrocellulose Millipore filters, and then run sequentially through 1 ml affinity columns containing either the control antibody (Hermes-1) or MECA-367 (2 mg antibody ml^{-1} CNBr-activated Sepharose 4B). Columns were washed extensively (0.1% Nonidet P-40, 500 mM NaCl, 5 mM EDTA, 0.02% NaN_3 and 1 mM PMSF in 50 mM Tris-HCl, pH 7.4) and eluted (0.1% Nonidet P-40, 500 mM NaCl, 0.02% NaN_3 , 1 mM PMSF in 0.2 M acetic acid). Eluted material was neutralized by adding 1 M Tris-HCl, pH 7.4, and concentrated in a Centricon-10 microconcentrator (Amicon). *a*, Immunoblot analysis of the MECA-367 antigen. Material eluted from each column was applied to an 8% SDS-PAGE gel, run under reducing conditions according to the method of Laemmli³⁸, and transferred to nitrocellulose with a BioRad transblot apparatus. Duplicate filters were blocked with 10% horse serum in TBST (10 mM Tris-HCl, pH 7.4, 150 mM NaCl and 0.05% Tween-20); incubated with either MECA-367, or control antibody (Hermes-1) at 100 $\mu g\ ml^{-1}$; washed in TBST; incubated with alkaline phosphatase-conjugated goat anti-rat immunoglobulin (Sigma A-9654) diluted 1:200; and washed again before addition of substrate solution (described in Promega λ gt11 screening kit). All incubations and washes were for 30 min in TBST. *b*, Analysis of iodinated MECA-367 antigen. The antigen was affinity isolated as above and iodinated by using Enzymobeads (BioRad, per manufacturer's directions). The labelled antigen was adsorbed and acid eluted from specific or control antibody-conjugated Sepharose 4B beads, under the conditions described above for initial affinity isolation, and electrophoresed as described³⁸ in a 10% SDS-PAGE gel.

specific antigen and related tissue-specific determinants in other sites 'vascular addressins'.

Tissue-specific anti-HEV monoclonal antibodies

Wistar rats were immunized with endothelial cells isolated from BALB/c mesenteric and peripheral lymph nodes for the production of hybridomas. Immunohistologic screening yielded two mAbs, MECA-89 and MECA-367, that intensely stained all HEVs in the mucosal lymphoid organs, the Peyer's patches, but failed to react with most HEVs in peripheral (axillary, brachial, popliteal and inguinal) lymph nodes (Fig. 1*a* and *b*; because

the staining properties of the two mAbs are indistinguishable, only sections stained with MECA-367 are shown). HEVs in some individual peripheral lymph nodes do exhibit focal positivity, either a faint diffuse staining or, in some rare instances, moderate staining of a few high endothelial cells (HECs) or of most HECs of individual HEVs. In immunohistologic studies, the cytoplasm of mucosal HECs appears intensely stained, and there is an impression of increased antigen at the luminal versus basal surface of the high endothelium. Immunofluorescence staining of isolated viable HECs in suspension confirms that both MECA-89 and MECA-367 antibodies define antigenic determinants that are expressed on the cell surface (see Fig. 1 legend).

HEVs in the mesenteric lymph node bind both peripheral lymph node-specific and Peyer's patch-specific HEV-binding lymphomas¹⁶, and are therefore thought to express both peripheral lymph node- and mucosa-associated endothelial cell determinants for lymphocyte recognition. Immunohistologic examination of mesenteric lymph nodes revealed that most HEVs in this tissue stain intensely with MECA-89 and MECA-367. Some HEVs or segments of HEVs, however, are only partially stained giving a mottled appearance; and others do not stain detectably (Fig. 1*c* and *d*). These results suggest clonal or microregional regulation of expression of the MECA-89 and MECA-367 antigen(s) by high endothelial cells. Mucosa-derived factors arriving in the mesenteric lymph node via the intestinal lymph may influence expression of these markers.

Immunohistologic examination of lamina propria in both small and large intestines revealed venules that stained with MECA-89 and MECA-367 (Fig. 1*e*). The intensity of staining was generally less than that of well developed HEVs in Peyer's patches, perhaps corresponding to the lower levels of lymphocyte traffic into the lamina propria. In serial sections, the MECA-367 positive lamina propria vessels also stained with a mAb, MECA-325 (ref. 23), that has been shown to identify lamina propria venules supporting the extravasation of transfused radiolabelled lymphocytes²⁴. Thus, the venules stained with MECA-89 or MECA-367 are likely to be involved in lymphocyte traffic to the mucosal lamina propria. The lactating mammary gland, another mucosal tissue known to support the extravasation of mucosa-specific lymphoblast populations such as IgA plasma cell precursors²⁵, also contains small vessels displaying the MECA-89 and MECA-367 antigenic determinants (not shown). In contrast, vessels in complete Freund's adjuvant-induced granulomas in a non-mucosal site, the skin, failed to stain with either antibody.

A 58–66K protein defined

The MECA-367 antigen was affinity isolated from detergent-solubilized mesenteric lymph node stroma, and characterized either by Western blot analysis (Fig. 2*a*) or by iodination, immunoprecipitation, and SDS-PAGE (Fig. 2*b*). In Western blots, a diffuse band with a relative molecular mass (M_r) in the range of 62K was detected by MECA-367 (Fig. 2*a*, lane 1) but not by control antibody (lane 2). Similarly, the iodinated, immunoprecipitated antigen ran as a single major band of about 58–66K under reducing conditions (Fig. 2*b*, lane 1). It migrated slightly more rapidly (M_r of 54–62K) under nonreducing conditions (lane 4). This shift in electrophoretic mobility with reduction may be due to internal disulphide bonding. Interestingly, the MECA-367 antigen was also bound by MECA-89 (lanes 2 and 5), but not by the control antibody (lanes 3 and 6), indicating that a single molecular species bears both the MECA-89 and MECA-367 epitopes. The epitopes are distinct, since MECA-89 does not cross-block immunohistologic staining of HEVs with MECA-367, or vice versa (see Fig. 1 legend).

Additional Western analyses (data not presented) showed that the passage of mesenteric node stromal lysate through a MECA-367 affinity column removed the MECA-89 antigen. Thus all MECA-89-reactive species appear to be recognized by

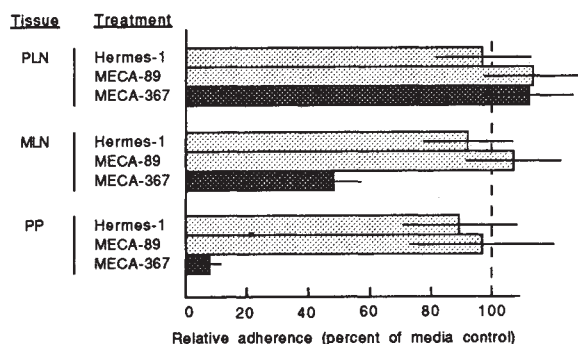


Fig. 3 MECA-367 inhibits lymphocyte binding to Peyer's patch and mesenteric lymph node HEVs. Results are presented as percent of media control ($\pm$ s.e.m.). PLN, peripheral (brachial, axillary, and inguinal) lymph nodes. MLN, mesenteric lymph node. PP, Peyer's patches.

Methods. *In vitro* lymphocyte/HEV binding assay. Our modification of the *in vitro* model of lymphocyte binding to HEVs in frozen sections has been described²⁷. Briefly, MECA-367 or MECA-89 were incubated at $100 \mu\text{g ml}^{-1}$ on $12 \mu\text{m}$ thick, freshly cut, unfixed tissue sections for 30 min at 7°C . The antibody-containing medium was decanted, and $1-3 \times 10^6$ mesenteric node lymphocytes were applied in $100 \mu\text{l}$ of RPMI 1640 medium containing 20 mM HEPES and 5% newborn calf serum (NCS). After incubation for 30 min with mild rotation at 7°C , the medium was decanted and sections were fixed in 1.5% glutaraldehyde in PBS. Lymphocyte binding to HEVs was assessed microscopically under darkfield illumination, and the mean number of bound cells per HEV was determined.

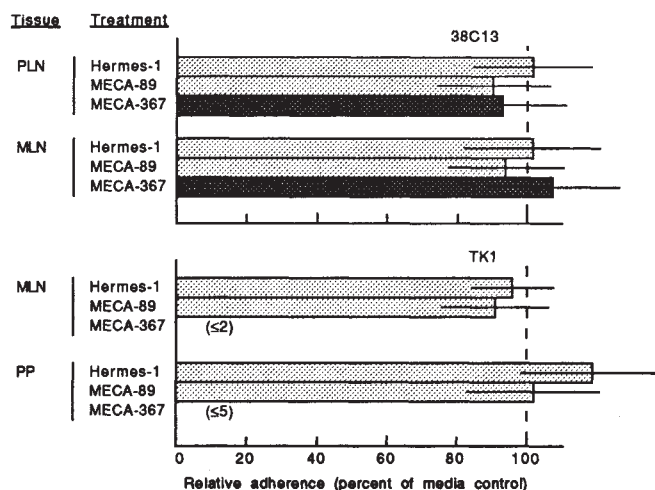


Fig. 4 Selective effect of MECA-367 on lymphoma cell binding to HEVs via mucosa-specific homing receptors. MECA-367, but not MECA-89 or control antibody, completely blocks the binding of a mucosal HEV-specific lymphoma (TK1) to Peyer's patch- and mesenteric node-HEVs. In contrast, MECA-367 has no effect on the binding of 38C13, a lymphoma that binds to peripheral lymph node and mesenteric node-HEVs via peripheral lymph node homing receptors, even when the HEVs bound (that is, in the mesenteric lymph node) are predominantly MECA-367 positive. Section pre-treatment and binding assay were conducted as described in Fig. 3. As no TK1 cells were seen binding to MECA-367-treated HEVs, the upper limits of 95% confidence based on the binomial distribution are given in parentheses.

MECA-367. It remains possible that a subset of the MECA-367-defined 58-66K species lacks the MECA-89 epitope.

Inhibition of lymphocyte-HEV binding

We next asked whether MECA-89 or MECA-367 could interfere with lymphocyte binding to HEVs in the Stamper-Woodruff *in vitro* assay^{26,27}. Frozen sections of BALB/c peripheral lymph nodes, mesenteric lymph nodes and Peyer's patches were incubated with saturating levels of MECA-89, MECA-367 or an isotype-matched control antibody, and the ability of the treated HEVs to bind normal mesenteric lymph node lymphocytes was determined (Fig. 3). Treatment with MECA-367 resulted in a 92% reduction in the ability of Peyer's patch HEVs to bind lymphocytes. MECA-367 also inhibited lymphocyte binding to mesenteric lymph node HEVs by 52%, consistent with the expression of both mucosal and peripheral lymph node endothelial cell determinants for lymphocyte recognition by these vessels (as discussed above). In contrast, the antibody had no detectable effect on lymphocyte binding to peripheral lymph node HEVs. Also, pretreatment of HEVs with MECA-367, followed by extensive washing, effectively inhibited binding; whereas pretreatment of lymphocytes did not alter their HEV-binding properties, ruling out an effect of the mAb on the lymphocyte. Neither the control antibody nor MECA-89 altered lymphocyte binding to the HEVs of any tissue.

The majority of normal lymphocytes are thought to express homing receptors for both lymph node and mucosal HEVs^{15,28}. To determine the effect of MECA-367 on HEV binding by cells that bind to HEVs in a tissue-specific manner, we employed two well-characterized lymphoma cell lines, TK1 and 38C13 (refs 16 and 19). MECA-367 completely blocked the binding of the mucosal HEV-binding cell line, TK1, to the HEVs in Peyer's patches and in mesenteric lymph nodes (Fig. 4). By contrast, it had no effect on the binding of the peripheral lymph node-specific cell line, 38C13, to peripheral or to mesenteric lymph node HEVs. Examination of subserial sections of mesenteric lymph nodes either incubated with 38C13 cells or stained

immunohistologically with MECA-367 confirmed that 38C13 cells bound well to vessel segments that were intensely MECA-367⁺. Thus, MECA-367 interferes exclusively with lymphocyte recognition of HEVs via a mucosa-specific recognition system, and has no effect on binding by a peripheral lymph node-associated system, even when the HEVs involved are intensely MECA-367⁺. These findings provide further evidence against a non-specific effect of MECA-367 on cell binding to MECA-367 positive HEVs.

Blockade of lymphocyte homing

To determine if the MECA-367 antigen was also required for lymphocyte extravasation under physiologic conditions *in vivo*, we labelled normal mesenteric lymph node cells with ^{51}Cr , and transfused them intravenously into mice that had been injected four hours previously with MECA-367, with control antibodies, or with media alone (Fig. 5). One hour after the administration of labelled cells, animals were sacrificed, and lymphocyte localization was determined by quantitating the counts per minute (c.p.m.) in various organs. Treatment with MECA-367 resulted in a nearly complete (97%) blockade of lymphocyte migration to Peyer's patches. In keeping with the *in vitro* results, the antibody caused a partial (37%) reduction in lymphocyte localization to the mesenteric lymph nodes, and had no significant effect on localization to peripheral lymph nodes. Control antibodies, whether against a distinct endothelial cell surface antigen or against an irrelevant antigen, had no significant effect on localization. Histologic examination of lymphoid tissues following administration of MECA-367 or control antibodies revealed no morphologic changes in HEVs. These results indicate that the MECA-367 antigen is specifically required for lymphocyte extravasation into organized mucosal lymphoid tissues *in vivo*.

Discussion

The two mAbs, MECA-89 and MECA-367, define a 58-66K endothelial cell protein antigen expressed by vessels at mucosal

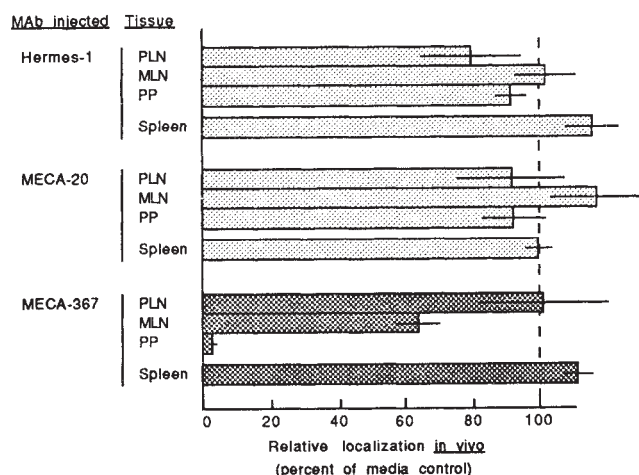


Fig. 5 Organ-specific blockage of lymphocyte homing *in vivo*. MECA-367, but not control antibodies, selectively inhibits lymphocyte extravasation into mucosal lymphoid organs. Results are expressed as per cent of cell localization observed in the media control animals ($\pm$ s.e.m.).

Methods. *In vitro* labelling of lymphocytes with ^{51}Cr : normal mesenteric node lymphocytes from 6–10 week old BALB/c mice were labelled by incubation of $1\text{--}2.5 \times 10^7$ cells per ml for 1 h at 37°C in DME medium supplemented with 20 mM HEPES, 5% fetal calf serum (FCS), and $100 \mu\text{Ci ml}^{-1}$ sodium chromate ($\text{Na}_2^{51}\text{CrO}_4$; New England Nuclear). After labelling, the cells were centrifuged through a layer of FCS, and washed twice with HBSS. *In vivo* homing: Recipient mice were injected intravenously (i.v.) either with HBSS alone, or with 1 mg of MECA-367, MECA-20 or Hermes-1 in HBSS. After four hours, 2.5×10^7 chromium labelled cells were delivered i.v., and one hour later the animals were sacrificed. Lymphocyte localization was determined by quantitating the c.p.m. in the indicated organs. There were four recipients for each antibody treatment.

sites of lymphocyte extravasation. MECA-367 inhibits the binding of normal and neoplastic lymphocytes to HEVs in the mucosa-associated lymphoid organs, Peyer's patches, but has no detectable effect on lymphocyte binding to HEVs in peripheral (non-mucosal) lymph nodes. This inhibition of binding to mucosal HEVs appears to reflect specific interference with recognition signals, rather than non-specific effects, in that (1) MECA-89, which reacts with the same vessels as MECA-367 and binds to the isolated MECA-367 antigen, has no effect on lymphocyte binding; and (2) MECA-367 fails to inhibit the ability of MECA-367⁺ mesenteric lymph node HEVs (which express both peripheral lymph node-associated and mucosa-associated binding specificities) to interact with lymphocytes via peripheral lymph node-specific homing receptors. In *in vivo* studies, intravenous administration of MECA-367 effectively blocks the homing of normal lymphocytes to mucosa-associated lymphoid tissues. Taken together, these data indicate that the MECA-367 antigen is involved in the recognition of mucosal HEVs by normal lymphocytes and transformed lymphoid cell lines. It is likely that the antigen either functions as an endothelial cell-surface ligand for mucosa-specific lymphocyte homing receptors, or that it is intimately associated both physically and functionally with such ligands. Formal proof of the recognition function of the molecule defined by MECA-89 and MECA-367 will require the demonstration that, as an isolated protein, it is capable of binding either lymphocytes or isolated lymphocyte homing receptors with appropriate specificity.

Molecules serving to convey positional information have been hypothesized to explain cellular interactions and positioning in a variety of systems, for example during morphogenesis and neurogenesis. The mucosa-specific vascular antigen defined here appears to represent such a molecule, in that it serves as a positional marker or address signal that is recognized by another

cell type. We propose to call this and related tissue-specific endothelial cell antigens involved in lymphocyte homing 'vascular addressins', where addressins are defined as molecules (1) that are expressed in an organ- or tissue-selective manner by a cellular or extracellular element of wider distribution; (2) that have functionally related counterparts expressed by the same cellular or extracellular element in other tissue sites; and (3) whose apparent role is to signal the position or mark the address of such cellular or extracellular components for the purpose of directing cell-cell interactions or cellular positioning in the body. The requirement for functionally related but cognitively distinct addressins in different tissues distinguishes addressins from cell adhesion molecules^{29,30}, and from topographically distributed markers that may define cell position in the context of a molecular gradient³¹.

The expression of the MECA-367 antigen on small vessels in both the lamina propria and the mammary gland suggests that this molecule functions as a common recognition element for lymphocyte extravasation in diverse mucosal sites, including both organized lymphoid and extralymphoid tissues. Indeed, preliminary experiments indicate that MECA-367 significantly inhibits the migration of mesenteric node (gut-homing) immunoblasts to the intestinal lamina propria. Such a common mucosal addressin, in combination with expression of mucosal homing receptors by mucosal effector lymphoid populations^{8–14,32}, may explain the unification of immune responses in widely separated mucosal tissues such as the intestines and the mammary gland into a 'common mucosal immune system'³³.

Evidence exists that carbohydrate on HECs plays an important role in lymphocyte recognition of lymph node and Peyer's patch HEVs^{34–36}, and it has been suggested that lymphocyte homing receptors are mammalian lectins binding tissue-specific carbohydrate (or carbohydrate-protein combined) determinants on endothelial cells^{34,35}. Thus, it will be of interest to examine the structure and functional role of carbohydrate elements on the 58–66K antigen described in this article.

Three independent lymphocyte-endothelial recognition systems are known, directing lymphocyte traffic to peripheral lymph nodes, to mucosal lymphoid organs, and to inflamed synovium in joints^{16–18}. The studies presented here show that at least two of these tissue-specific recognition systems involve antigenically distinct endothelial cell-surface molecules: MECA-89 and MECA-367 define antigenic determinants on the mucosal addressin on Peyer's patch HEVs, and fail to identify its functionally defined counterpart on peripheral lymph node HEVs. Additional tissue-specific vascular addressins may operate in other sites to direct lymphocyte extravasation to many organs in the body. Analyses of the nature and regulation of these endothelial-cell addressins may lead to a better understanding not only of lymphocyte trafficking, but also of the role of vascular differentiation in the organ-specific metastasis of neoplasms, and of fundamental cellular mechanisms for reporting cellular positioning and tissue associations in the body.

We thank J. E. Streeter for preparation of the illustrations, S. Grossman for secretarial assistance, and the NIH for support. P.R.S. was supported by a fellowship from the Bank of America Giannini Foundation and is currently a Fellow of California Division of the American Cancer Society, E.L.B. is a Fellow of the American Cancer Society, and E.C.B. is an Established Investigator of the American Heart Association.

Received 24 July; accepted 19 November 1987.

- Alby, L. & Auerbach, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5739–5743 (1984).
- Auerbach, R. *et al. Cancer Res.* **47**, 1492–1596 (1987).
- Bargatze, R. F., Wu, N. W., Weissman, I. L. & Butcher, E. C. *J. exp. Med.* **166**, 1125–1131 (1987).
- Gordon, S. *J. Cell Sci. Suppl.* **4**, 267–286 (1986).
- Lewinson, D. M., Bargatze, R. F. & Butcher, E. C. *J. Immun.* **138**, 4313–4321 (1987).
- Gowans, J. L. & Knight, E. J. *Proc. R. Soc. B159*, 257–282 (1964).
- Marchesi, V. T. & Gowans, J. L. *Proc. R. Soc. B159*, 283–290 (1964).
- Guy-Grand, D., Griscelli, C. & Vassalli, P. *Eur. J. Immun.* **4**, 435–443 (1974).
- McWilliams, M., Phillips-Quagliata, J. M. & Lamm, M. E. *J. exp. Med.* **145**, 866–875 (1977).
- Rose, M. L., Parrott, D. M. V. & Bruce, R. G. *Immunology* **35**, 415–423 (1978).

11. Hall, J. G., Hopkins, J. & Orlans, E. *Eur. J. Immun.* **7**, 30–37 (1979).
12. Cahill, R. N. P., Poskitt, D. C., Frost, H. and Trnka, A. *J. exp. Med.* **145**, 420–428 (1977).
13. Smith, M. E., Martin, A. F. & Ford, W. L. *Monogr. Allergy* **16**, 203–232 (1980).
14. Chin, W. & Hay, J. B. *Gastroenterology* **79**, 1231–1242 (1980).
15. Butcher, E. C. *Curr. Topics Microbiol. Immun.* **128**, 85–122 (1986).
16. Butcher, E. C., Scollay, R. G. & Weissman, I. L. *Eur. J. Immun.* **10**, 556–561 (1980).
17. Chin, Y. H., Rasmussen, R., Cakiroglu, A. G. & Woodruff, J. J. *J. Immun.* **133**, 2961–2965 (1984).
18. Jalkanen, S., Steere, A. C., Fox, R. I. & Butcher, E. C. *Science* **233**, 556–558 (1986).
19. Gallatin, W. M., Weissman, I. L. & Butcher, E. C. *Nature* **303**, 30–34 (1983).
20. Chin, Y. H., Rasmussen, R. R., Woodruff, J. J. & Easton, T. G. *J. Immun.* **136**, 2556–2561 (1986).
21. Jalkanen, S., Bargatzte, R. F., Herron, L. & Butcher, E. C. *Eur. J. Immun.* **16**, 1195–1202 (1986).
22. Jalkanen, S. T., Bargatzte, R. F., de los Toyos, J. & Butcher, E. C. *J. Cell Biol.* **105**, 983–990 (1987).
23. Duijvestijn, A. M., Kerkhove, M., Bargatzte, R. F. & Butcher, E. C. *J. Immun.* **138**, 713–719 (1987).
24. Jeurissen, S. H. M., Duijvestijn, A. M., Sontag, Y. & Kraal, G. *Immunology* **62**, 273–277 (1987).

25. Roux, M. E., McWilliams, M., Phillips-Quagliata, J. M., Weisz-Carrington, P. & Lamm, M. E. *J. exp. Med.* **146**, 1311 (1977).
26. Stamper, H. B. Jr & Woodruff, J. J. *J. exp. Med.* **144**, 828–833 (1976).
27. Butcher, E. C., Scollay, R. G. & Weissman, I. L. *J. Immun.* **123**, 1996–2003 (1979).
28. Stevens, S. K., Weissman, I. L. & Butcher, E. C. *J. Immun.* **128**, 844–851 (1982).
29. Edelman, G. M. *Science* **219**, 450–457 (1983).
30. Springer, T. A., Dustin, M. L., Kishimoto, T. K. and Marlin, S. D. *A. Rev. Immun.* **5**, 223–252 (1987).
31. Trisler, G. D., Schneider, M. D. & Nirenberg, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2145–2149 (1981).
32. Schmitz, M., Nunez, D. & Butcher, E. C. *Gastroenterology* (in the press).
33. McDermott, M. R. & Bienenstock, J. *J. Immun.* **122**, 1892–1898 (1979).
34. Stoolman, L. M. & Rosen, S. D. *J. Cell Biol.* **96**, 722–729 (1983).
35. Stoolman, L. M., Tenforde, T. & Rosen, S. D. *J. Cell Biol.* **99**, 1535–1540 (1984).
36. Rosen, S. D., Singer, M. S. & Yednock, T. A. *Science* **228**, 1005–1007 (1985).
37. Duijvestijn, A. M., Schreiber, A. B. & Butcher, E. C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9114–9118 (1986).
38. Laemmli, V. K. *Nature* **227**, 680–685 (1970).

LETTERS TO NATURE

A test of the massive binary black hole hypothesis: Arp 102B

J. P. Halpern* & Alexei V. Filippenko†

* Columbia Astrophysics Laboratory, 538 West 120th Street, New York, New York 10027, USA

† Astronomy Department, University of California, Berkeley, California 94720, USA

The emission-line spectra of several active galactic nuclei (AGNs) have broad peaks which are significantly displaced in velocity with respect to the host galaxy. An interpretation of this effect in terms of orbital motion of a binary black hole predicts periods of a few centuries^{1,2}. Here we point out that recent measurements of the masses and sizes of many low-luminosity AGNs imply orbital periods much shorter than this. In particular, we find that the elliptical galaxy Arp 102B is the most likely candidate for observation of radial velocity variations; its period is expected to be ~ 3 yr. We have monitored the H α line profile of Arp 102B for 5 yr without detecting any change in velocity, and thus find that a rather restrictive observational test of the massive binary black hole hypothesis already exists, albeit for this one object.

Gaskell's estimate of a period of a few centuries¹ is based upon the reasonable assumption that the binary separation has to be at least as large as the typical radius of the broad-line region (0.1–1 pc) for separate emission-line peaks to be discernible. Kepler's third law yields $P = 550 R_{18}^{3/2} M_9^{-1/2}$ yr, where R_{18} is the binary separation in units of 10^{18} cm and M_9 is the sum of the black hole masses in units of $10^9 M_\odot$. The expected velocity, assuming equal masses, is $v = 1,800 M_9^{1/2} R_{18}^{-1/2}$ km s⁻¹, in agreement, for a pair of $10^9 M_\odot$ black holes, with observed displacements of 2,000–3,000 km s⁻¹. Although these estimates for M and R are representative of luminous quasars, several of the objects in which the displaced-peak phenomenon has been observed are better characterized in terms of absolute luminosity as Seyfert or radio galaxies. The well-known example³ 3C390.3 has redshift $z = 0.057$, $M_{\text{abs}} = -22.3$, and X-ray luminosity⁴ $\sim 3 \times 10^{44}$ erg s⁻¹. A cross-correlation analysis of the continuum and emission-line light curves indicates that the broad-line region in 3C390.3 is only 10^{17} cm in radius⁵, so an orbital period could be as small as 15–20 yr. Recent observations of X-ray variability in Seyfert galaxies⁶ have provided estimates of the black hole size via light-travel time arguments, which seem to indicate that the mass is proportional to the X-ray luminosity⁷. An equivalent statement is that the luminosity is a fixed fraction of the Eddington value; in this case the bolometric luminosity is $10^{-2.0 \pm 0.5} L_{\text{Edd}}$ (ref. 7). Thus at fixed orbital velocity, the lower luminosity object is likely to have the shorter period. Several Seyferts have been seen to vary significantly in X-rays on time scales of an hour^{8–12}, which implies $M \leq 10^8 M_\odot$. In addition, correlated continuum and line variations which have been seen in some Seyfert galaxies indicate that the broad-line region may be of the order of 1 light

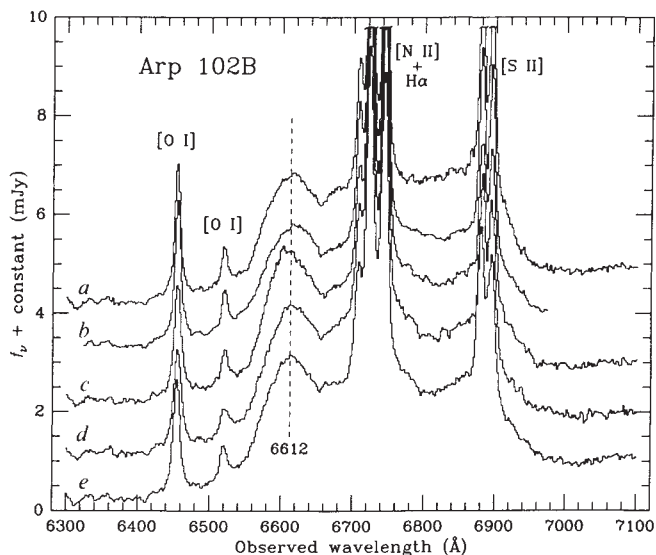


Fig. 1 CCD spectra of the nucleus of Arp 102B obtained on (a) 19 June 1983, (b) 28 June 1985, (c) 13 July 1986 and (d) 4 May 1987. Spectrum (e) is an average of two observations made on 8 and 12 August 1987, which are indistinguishable within noise limits. The ordinate scale refers to spectrum (c). All other spectra were scaled to the average flux of (c) and are plotted with additive offsets of 1 or 2 mJy. The vertical dashed line is at 6,612 Å, the average wavelength of the prominent blueshifted peak in the profile of H α .

week in size¹³, which would permit still shorter orbital periods to be detected.

Arp 102B ($M_{\text{abs}} = -21$, $z = 0.0244$), an elliptical AGN of modest radio¹⁴ and X-ray¹⁵ luminosity, is the best candidate we know of for a test of the binary hypothesis. Its broad H α line profile^{16,17} is very similar to that of 3C390.3, with a prominent peak blueshifted by 5,000 km s⁻¹ from the rest frame of the galaxy as defined by either the narrow emission lines or the stellar continuum. The X-ray luminosity¹⁵ is only 8.6×10^{42} erg s⁻¹, the emission-line luminosities scale appropriately¹⁶, and the optical continuum is dominated by starlight. Although no observations for X-ray variability have been made, it is reasonable to assume that the mass of the black hole in Arp 102B is $\leq 10^8 M_\odot$, similar to that derived⁷ for virtually all Seyfert galaxies having $L_x \leq 10^{43}$ erg s⁻¹. Given a binary with equal masses, Kepler's third law can be recast as $P \leq 2.7 M_8 (v \sin i)^{-3}_{5,000}$ yr, where i is the inclination angle of the orbit with respect to the plane of the sky, and $(v \sin i)_{5,000}$ is the observed radial velocity in units of 5,000 km s⁻¹. The small inferred mass, together with the relatively large radial velocity (which is of course only a lower limit to the orbital velocity), conspire to bring the predicted period within the span of a few observing seasons.

Table 1 Spectroscopy of Arp 102B

UT date	Telescope or reference	Resolution (Å)	Wavelength of displaced peak (Å)
6 Jun 1982	Ref. 16	10	6,608 ± 7
19 Jun 1983	Palomar 5-m	2	6,615 ± 3
15 Sep 1983	Palomar 5-m	15	6,610 ± 5
28 Jun 1985	Palomar 5-m	2	6,616 ± 4
13 Jul 1986	Lick 3-m	7	6,606 ± 4
4 May 1987	Lick 3-m	7	6,613 ± 3
8 Aug 1987	Lick 3-m	7	6,615 ± 4
12 Aug 1987	Lick 3-m	7	6,612 ± 3

In fact, we have been recording high-quality spectra of the H α line of Arp 102B since June 1983, using CCD (charged couple device) spectrographs on the 5-m Hale telescope at Palomar Observatory and the 3-m Shane reflector at Lick Observatory. Most of the observations were made in photometric conditions. The slit width was 2" and the atmospheric seeing was 1"–1.5". Normal procedures¹⁸ were used to calibrate and analyse the spectra. In addition, small errors in the wavelength scale were eliminated with the help of atmospheric emission lines and narrow emission lines in Arp 102B. An accurate redshift, derived from the narrow lines in several of the best spectra, is 0.02438 ± 0.00006 .

The broad component of H α (Fig. 1) departs dramatically from the gaussian or logarithmic profile normally seen in AGNs. Although the profile varies with time, in all cases there is a well-isolated hump on the blue side of the line. A less prominent hump is also visible on the red side, but we will not analyse it here because the narrow [S II] emission lines contaminate it.

It is clear that the wavelength of the blue hump remains virtually constant. Asymmetries in this hump make its exact wavelength difficult to define, but the difference between the peak and centroid is rarely more than a few ångströms, so an average of these quantities is adopted. Our measurements of the wavelength of the hump are listed in Table 1. The quoted errors are about one standard deviation. We also list a measurement from a published spectrum¹⁶ which extends the temporal coverage to 1982. The measurements are consistent with a constant wavelength of 6,612 Å, which translates to a velocity of 4,980 km s⁻¹ in the frame of the galaxy. The wavelengths display no systematic variation with time.

We have measured the velocity of the displaced broad-line peak to an accuracy of about 200 km s⁻¹, which is 4% of the observed velocity. In 5 years there have been no detectable changes in velocity at the 4% level; hence, we can place an upper limit of $2 \cos^{-1} 0.96 = 32^\circ$ on $\Delta\phi$, the change in phase angle of the hypothesized binary. This translates to a lower limit of 55 yr for an orbital period. Thus, we can claim confidently that either the mass is greater than $2 \times 10^9 M_\odot$, or Arp 102B is not a spectroscopic binary.

A similar analysis of possible evidence for a binary black hole in the Seyfert galaxy NGC5548 relied on a careful decomposition of the line profile into two components which are stationary in wavelength¹⁹. On this interpretation, the binary period of NGC5548 is at least 110 yr and the mass is greater than $4 \times 10^7 M_\odot$. Since this mass limit is not in conflict with observations or models of NGC5548, the data on this galaxy are not yet a test of Gaskell's hypothesis, but are consistent with it.

Of course, the weak link in this method is the estimation of the unknown mass from the X-ray luminosity, in the absence of variability data. Although the proportionality between mass and X-ray luminosity seems to be good to within a factor of three for type 1 Seyfert galaxies⁶, unusual elliptical galaxies such as Arp 102B have not been studied as carefully for X-ray variability. Ellipticals may well be underluminous at X-ray energies compared with Seyferts of the same central mass. In fact, the X-ray luminosity of 3C390.3 has declined steadily by more

than a factor of 10 in the past 15 yr (ref. 4), so it is difficult to use this quantity as a measure of the mass. The fact that most of the displaced-peak objects are radio loud¹ may cast doubt upon the adoption of results gleaned from a largely X-ray selected (mostly radio quiet) sample⁶. Nevertheless, future X-ray observations of Arp 102B could, in principle, settle the question sooner than continued optical monitoring. Any observation of substantial variations in less than 8 h would imply that $M \leq 10^9 M_\odot$, eliminating the spectroscopic binary hypothesis.

Even a binary period of 55 yr, which is barely consistent with the current observational data, poses an additional theoretical difficulty. The corresponding separation of $\sim 3 \times 10^{17}$ cm is comparable with the radius of 5×10^{17} cm within which gravitational radiation takes over as the dominant force in the dynamical evolution². Gravitational radiation shrinks this orbit on a time scale of 2×10^7 yr (ref. 2), much less the expected time of $\geq 10^9$ yr spent at larger separations during which the evolution is controlled by dynamical friction and gas accretion². Thus, it is somewhat unlikely that one of only a handful of binary black holes would be observed in a state of small separation. On the other hand, there could be many more binaries of larger separation, such as NGC5548, which have gone unnoticed, making the small separation in Arp 102B less improbable.

Arp 102B forms an ideal test for Gaskell's hypothesis that binary black holes are formed by capture during collisions between galactic nuclei, one of which is a dwarf elliptical²⁰. Although not a dwarf, Arp 102B is clearly an elliptical galaxy which has interacted with a disturbed spiral companion²¹. Only the elliptical currently has an active nucleus, so that the detection of a binary black hole would be strong evidence in favour of the capture theory. On the other hand, if continued monitoring of Arp 102B restricts the parameters of the binary model until it becomes implausible, another explanation of the unusual line profile will have to be found.

Perhaps models involving ordered motion in an accretion disk or jet should be considered. Some U Geminorum stars show double-peaked line profiles, but these tend to appear at small velocity, indicating an origin in the outer part of the disk^{22–24}. For example, U Gem²² had peaks separated by 1,100 km s⁻¹, but the full width of the line is 5,000 km s⁻¹. The best-fit disk model for U Gem has a ratio of inner to outer radius of ~ 0.04 . By contrast, the displaced peaks in Arp 102B and other AGNs occur near the maximum velocity of the profile. Models for line profiles from a rotating disk in quasars have been calculated with relativistic effects included^{25,26}. The line profile of Arp 102B is similar to disk models in which the ratio of inner to outer radius is ~ 0.3 . The slight asymmetries could be accounted for by relativistic effects. Oke²⁷ has also suggested that a disk could be responsible for the displaced peaks in 3C390.3.

Alternatively, the displaced peaks may be due to material entrained in a pair of jets emerging along the axis of the disk. If the jets are longer than the disk radius, then both blue and red peaks should be observed. If the jets are short and the disk is optically thick, then only the blue peak is expected to be visible. This is analogous to the interpretation of the X-ray iron line in SS433, which shows only the blueshifted component²⁸. In any case, the tendency of displaced peaks to be found in radio-loud objects is undoubtedly a clue to their origin, as is the unusually flat profile which characterizes the undisplaced portions of the lines in Arp 102B and 3C390.3.

The observations were carried out at Palomar and Lick Observatories. This work was supported by NSF (Lick Observatory), Cal Space (A.V.F.), and NASA (J.P.H.). J.P.H. thanks Kaiyou Chen for helpful discussions. This paper is contribution number 338 of the Columbia Astrophysics Laboratory.

Received 8 September; accepted 21 October 1987.

1. Gaskell, C. M. *Liege Astrophysical Colloq.* **24**, 473–477 (1983).
2. Begelman, M. C., Blandford, R. D. & Rees, M. J. *Nature* **287**, 307–309 (1980).
3. Osterbrock, D. E., Koski, A. T. & Phillips, M. M. *Astrophys. J.* **206**, 898–909 (1976).
4. Shaffer, R., Ward, M. & Barr, P. *Space Sci. Rev.* **40**, 637–641 (1983).

5. Clavel, J. & Wamsteker, W. *Astrophys. J.* **320**, L9-L14 (1987).
6. Barr, P. & Mushotzky, R. F. *Nature* **320**, 421-423 (1986).
7. Wandel, A. & Mushotzky, R. F. *Astrophys. J.* **306**, L61-L65 (1986).
8. Tennant, A. F. & Mushotzky, R. F. *Astrophys. J.* **264**, 92-104 (1983).
9. Lawrence, A., Watson, M. G., Pounds, K. A. & Elvis, M. *Mon. Not. R. astr. Soc.* **217**, 685-699 (1985).
10. Pounds, K. A., Turner, T. J. & Warwick, R. S. *Mon. Not. R. astr. Soc.* **221**, 7P-12P (1986).
11. McHardy, I. & Czerny, B. *Nature* **325**, 696-698 (1987).
12. Turner, T. J. *Mon. Not. R. astr. Soc.* **226**, 9P-13P (1987).
13. Gaskell, C. M. & Sparke, L. S. *Astrophys. J.* **305**, 175-186 (1986).
14. Puschell, J. J., Moore, R., Cohen, R. D., Owen, F. N. & Phillips, A. C. *Astr. J.* **91**, 751-754 (1986).
15. Biermann, P., Kronberg, P. P., Preuss, E., Schilizzi, R. T. & Shaffer, D. B. *Astrophys. J.* **250**, L49-L53 (1981).
16. Stauffer, J., Schild, R. & Keel, W. *Astrophys. J.* **270**, 465-470 (1983).
17. Halpern, J. P. & Oke, J. B. *Astrophys. J.* **301**, 753-758 (1986).
18. Filippenko, A. V. & Sargent, W. L. W. *Astrophys. J. Suppl.* **57**, 503-522 (1985).
19. Peterson, B. M., Korista, K. T. & Cota, S. A. *Astrophys. J.* **312**, L1-L4 (1987).
20. Gaskell, C. M. *Nature* **315**, 386 (1985).
21. Arp, H. C. *Atlas of Peculiar Galaxies* (California Institute of Technology, Pasadena, 1966).
22. Stover, R. J. *Astrophys. J.* **248**, 684-695 (1981).
23. Smak, J. *Acta astr.* **19**, 155-164 (1969).
24. Huang, S. *Astrophys. J.* **171**, 549-564 (1972).
25. Gerbal, D. & Pelat, D. *Astr. Astrophys.* **95**, 18-23 (1981).
26. Mathews, W. G. *Astrophys. J.* **258**, 425-433 (1982).
27. Oke, J. B. in *Superluminal Radio Sources* (eds Zensus, J. A. & Pearson, T. J.) 267-272 (Cambridge University Press, 1987).
28. Watson, M. G., Stewart, G. C., Brinkmann, W. & King, A. R. *Mon. Not. R. astr. Soc.* **222**, 261-271 (1986).

A possible 35-minute periodicity in the OJ 287 active galactic nucleus at 7-mm wavelength

W. M. Kinzel, R. L. Dickman & C. R. Predmore

Five College Radio Astronomy Observatory and Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts 01003, USA

BL Lacertids are a class of active galactic nuclei (AGNs) characterized in part by flat radio spectra and rapid flux variability on time scales of weeks to days, at centimetre to optical wavelengths¹. One such object, OJ 287, exhibits very short time-scale (minutes) nonperiodic variations in infrared and centimetre flux, as well as in optical polarization²⁻⁴. Over the past fifteen years, there have also been reports of short term (tens of minutes) periodic⁵⁻⁸ modulation of the emission from OJ 287 at a variety of wavelengths; these periodicities have proved elusive to confirm^{3,9}, perhaps because they are inherently sporadic. Because the presence of periodic flux variations strongly constrains the nature of the central source within an AGN, the report of periodic variability in OJ 287 with a stable 15.7-min period detected at multiple frequencies and using different telescopes⁸, prompted us to undertake observations of this object at 7-mm wavelength. In this letter, we report detection in February 1986 of a weak (4% amplitude), 35-min flux modulation of OJ 287, whose statistical significance we estimate to be ~97%. Nearly one year later, observations of the same source revealed no periodicities between 4 and 300 min with amplitudes greater than 2%.

Observations of five compact radio sources including OJ 287 were made during two days in February 1986 with the 14 m

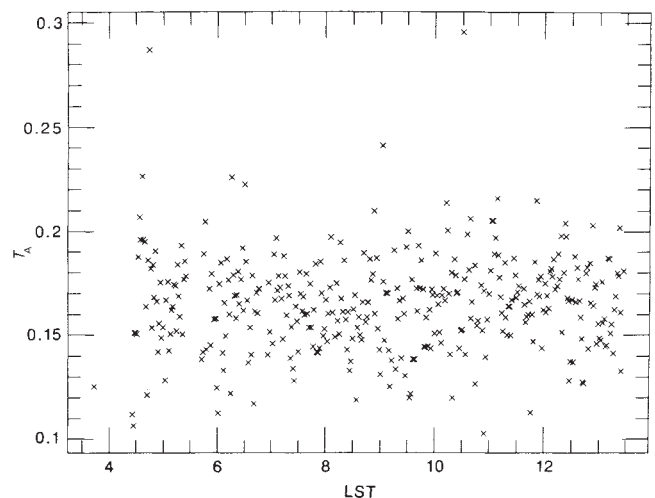


Fig. 1 Antenna temperatures for OJ287 measured on the morning of 1 Feb. 1987. The mean value is 0.165 ± 0.001 K. The typical error in a single measurement is 0.016 K.

FCRAO (Five College Radio Astronomy Observatory) radio telescope operating at a frequency of 42.6 GHz. Additional observations of OJ 287 alone were also made on 1987 January 27, for reasons described below. The cooled 7-mm receiver¹⁰ had an instantaneous bandwidth of approximately 350 MHz and was operated in double sideband mode with an intermediate frequency of 1.4 GHz. The system temperature was measured every few scans using an ambient-temperature absorbing vane¹¹. In order to achieve maximum baseline stability, all sources were observed by beam chopping at 15 Hz against blank sky displaced about 6 arc min in azimuth from the source position, and alternating the main and reference beams on source every 12 s. Each scan consisted of a pair of such cycles, giving a total integration time of 48 s. Because the FCRAO antenna has an altazimuth mounting, the approximately 15-s overhead time for each scan (the time required to move alternately between the source and reference position, an interval which depends upon the source elevation) varied as the sources moved across the sky. Combined with the calibration measurements carried out every few scans, this produced an irregularly-sampled data series for each source. The irregularity of the sampling reduces sidelobes in the spectral window which could cause spurious features in the periodogram analysis¹². Table 1 shows the sources and dates of observation.

The data for each source were analysed using the periodogram technique developed by Scargle¹³ and elucidated by Horne and Baliunas¹⁴. The method allows one to assign a significance level to peaks in the estimated power spectrum of an irregularly-sampled data stream, provided that the time-series can be regarded as consisting of uncorrelated gaussian noise. Because no observational data can be expected to satisfy completely such a stringent statistical prerequisite (the power spectrum of virtually any astronomical time series will always be slightly 'red', owing to the presence of long-term environmental and receiver

Table 1 Summary of observations

Date	Source	Periods examined (min)	N*	Period detected (min)	Amplitude (%)	p† (%)
1 Feb. 1986	OJ 287	3-300	356	35	4	<3
	BL Lac	2-20	43	—	<40	70
	3C84	3-120	141	—	<1	40
	SGR AW	3-50	67	—	<2	20
16 Feb. 1986	SGR AW	4-100	111	—	<1	70
	1921-29	4-40	40	—	<7	50
27 Jan. 1987	OJ 287	4-300	264	—	<2	75

* Number of data points in time series.

† Probability that largest peak in periodogram is due to gaussian noise.

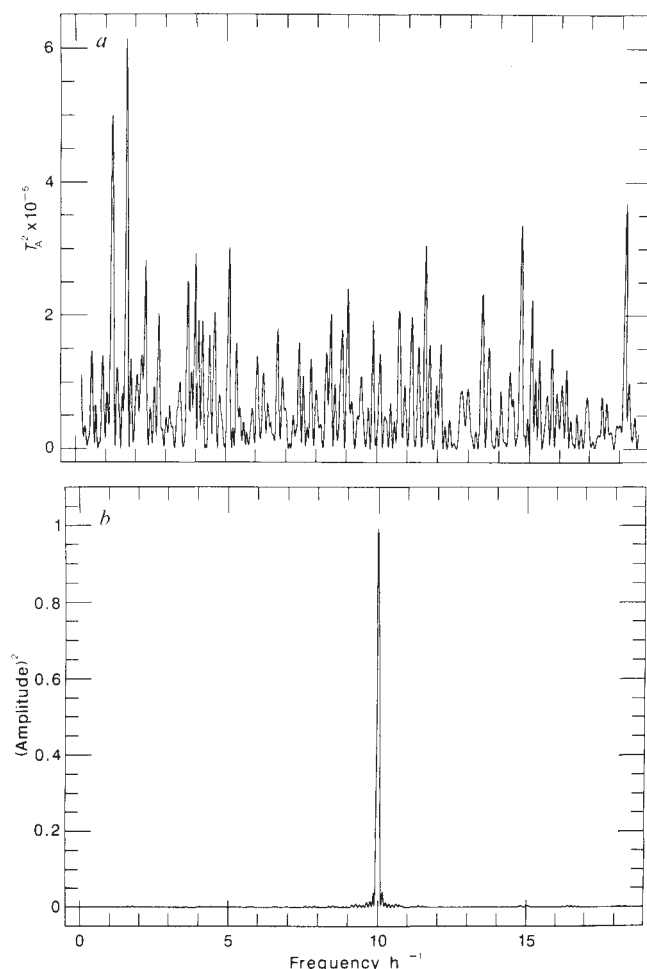


Fig. 2 Periodogram (*a*) of the data in Fig. 1. The vertical axis has been normalized so that the peak value is equal to the square of the amplitude of the corresponding sine wave embedded in the data. The pseudowindow (*b*) for the data sampling shown in Fig. 1, generated by calculating the periodogram of a sine wave with unit amplitude and a frequency of 10 h^{-1} using the sampling of Fig. 1 (ref. 13).

gain trends, for example), the significance levels calculated by this procedure should be viewed as approximate. Before calculating periodograms for the data, a least squares polynomial fit of up to second order was removed, if necessary, from each data set to reduce the effects of slowly-varying changes in the gain of the system. Periodograms were computed after removing the mean of each corrected data set, for sample frequencies between $2/T$ and $N/(2T)$, where T is the time over which the source was observed and N is the number of data points^{13,14}.

No statistically significant periodicities were observed in the data for any source except OJ 287 (Table 1). Figure 1 shows the measured data obtained for OJ 287 on the morning of 1 February 1986; only a mean value was removed from the data before the periodogram was calculated, since a second-order least-squares fit to the data resulted in linear and quadratic baseline coefficients of no statistical significance. The periodogram (Fig. 2*a*) shows a peak corresponding to a sine-wave amplitude of 4% and a period of 35 min; also shown (Fig. 2*b*) is the pseudowindow of the data sampling¹³. Within the statistical limitations noted above and because the data were not uniformly sampled, the formal probability that this spectral feature in the data occurs because of random gaussian noise is less than 3%¹⁴. There is also a second feature visible in Fig. 2, with a period of about 49 min, but further analysis (see ref. 14) shows it to be of no statistical significance, with a probability of about 40%

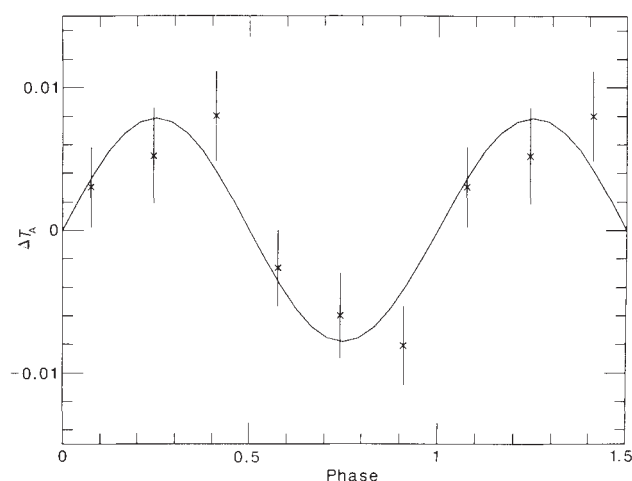


Fig. 3 Data folded into six bins using a period of 35 min. The plot shows the means from each bin, with error bars twice the error in the mean. Also shown is a least-squares fit sine wave with a period of 35 min.

that it is simply due to random fluctuations, and we ignore it hereafter.

The 35-min peak in the periodogram of Fig. 2*a* does not appear to be an artefact of the analysis method used. For example, we have also used a folding method¹⁵ which clearly indicates the presence of a 35-min period; the folded data, binned and averaged, is shown in Fig. 3.

The data of 1 February 1986 thus appear to contain a weak, periodic component having a rather high (formal) statistical significance. We have considered ways this periodicity might arise within the radio telescope system itself, and exclude them as follows. The 35-min period is long enough that an amplifier-related origin of the periodicity seems unlikely. This is supported by the stability of the system gain during the entire 1986 observing run, and by the fact that no other object observed at that time exhibited periodic behaviour (Table 1). Other possible systematic sources of spurious periodicities are periodic antenna encoder errors or scattering by the quasi-random radome

Table 2 Observed periodicities of OJ 287

Date(s)	Frequency	Period (min)	Amplitude (%)	Reference
16–21 Mar. 1972	‡	39.2	1	5
2 Feb. 7 Mar. 1973	‡	—	<0.15	9
9, 10 Mar. 1973	‡	40.0	~1	6
24 Apr. 1981	37 GHz	15.7	*	8
26 Feb. 1982	22 GHz	15.7	*	8
	§	16.3	*	8
	§	15.6	*	8
5–7 Mar. 1982	22 GHz	15.7	*	8
Jan.–Mar. 1983	§	22.8	3	7
		42.2	2	7
23 May 1983	5 GHz	—	<0.1	3
	22 GHz	15.7	0.5	8
	22 GHz	13.0	*	8
	37 GHz	15.7	*	8
	37 GHz	13.0	*	8
1 Feb. 1986	42 GHz	35.1	4	This work
20–24 Feb. 1986	8 GHz	—	<(0.7–3.)†	
27 Jan. 1987	42 GHz	—	<2	This work

* Amplitude not stated by authors.

† Amplitude limit depends upon day.

‡ Unfiltered optical photometer.

§ Optical photometer with B-band filter.

|| M. M. Komesaroff, J. A. Roberts & J. D. Murray (personal communication).

support frame; both, if present, might plausibly be associated with the altazimuth track of a particular source. To test this possibility, we observed OJ 287 approximately one year later under almost identical observing conditions and using identical observing procedures. During the year separating the two observations, the only change made to the FCRAO telescope system was the installation of a more sensitive continuum back-end. Periodogram analysis of the data resulted in an upper limit of less than 2% for the amplitude of any modulation (Table 1). This result strongly suggests that the periodic modulation of the flux from OJ 287 was not caused by the track of the source through the radome and is unlikely to be due to periodic pointing errors arising from the telescope drives.

It is of course still possible that the peak in the periodogram of Fig. 2 is due to an unlikely combination of random variations, or to systematic effects of which we are ignorant. The explanation most simply consistent with the known facts, however, is that a periodic modulation of the flux from OJ 287 occurred on 1 February 1986, and that this modulation originated outside the radome of the FCRAO telescope.

By itself, the 35-min periodicity does not provide new physical insight into the modulation mechanism. For example, although the existence of a 35-min periodic component in OJ 287 is consistent with the suggestion that periodic flux modulation can arise from the orbital motion of hot spots in an accretion disk about a massive central object¹⁶⁻¹⁸, the period found by us is longer than that claimed by other groups^{7,8}, and therefore cannot better constrain the central mass powering the accretion disk in this type of model. However, an interesting issue is raised when the results of periodicity studies over the past fifteen years are examined in aggregate. These are shown in Table 2. At least two points emerge from inspection of this table. First, if all the

observations in the table are given equal credence, it is clear that we must be dealing with a phenomenon that is sporadic and whose period changes with time. This is consistent with the picture of a hot spot in an accretion disk, since viscous processes might be expected to cause both the evolution of the spot's orbital frequency as well as its ultimate dissolution. Second, two of the tabulated null results, which were made nearly coincident with detected periodicities, were obtained at significantly lower frequencies (8 and 5 GHz) than the lowest frequencies (43 and 22 GHz) at which the periodic flux modulations were reported. This suggests that an opacity effect cuts off or washes out periodic flux variations at low frequencies in this source. In view of these issues, we feel it is imperative that future attempts to observe flux periodicities in this source be made nearly as simultaneously as possible, by more than one telescope operating at the same frequency, at K-band, or higher.

Received 17 August; accepted 20 October 1987.

1. Wiita, P. J. *Phys. Rep.* **123**, 117-213 (1985).
2. Wolstencroft, R. D., Gilmore, G. & Williams, P. M. *Mon. Not. R. astr. Soc.* **201**, 479-485 (1982).
3. Dreher, J. W., Roberts, D. H. & Lehar, J. *Nature* **320**, 239-242 (1986).
4. Kulshrestha, A. K., Joshi, U. C. & Deshpande, M. R. *Nature* **311**, 733-734 (1984).
5. Visvanathan, N. & Elliot, J. L. *Astrophys. J.* **179**, 721-730 (1973).
6. Frolich, A., Goldsmith, S. & Weistrop, D. *Mon. Not. R. astr. Soc.* **168**, 417-425 (1974).
7. Carrasco, L., Dultzin-Hacyan, D. & Cruz-Gonzalez, I. *Nature* **314**, 146-148 (1985).
8. Valtaoja, E. et al. *Nature* **314**, 148-149 (1985).
9. Kiplinger, A. J. *Astrophys. J.* **191**, L109-L110 (1974).
10. Predmore, C. R. *Five College Radio Astronomy Observatory Report No. 310* (1987).
11. Penzias, A. A. & Burrus, C. A. *A. Rev. Astr. Astrophys.* **11**, 51-72 (1973).
12. Deeming, T. J. *Astrophys. Space Sci.* **36**, 137-158 (1975).
13. Scargle, J. D. *Astrophys. J.* **263**, 835-853 (1982).
14. Horne, J. H. & Baliunas, S. L. *Astrophys. J.* **302**, 757-763 (1986).
15. Jurkevich, I. *Astrophys. Space Sci.* **13**, 154-167 (1971).
16. Sunyaev, R. A. *Sov. Astr.* **16**, 941-943 (1973).
17. Elliot, J. L. & Shapiro, S. L. *Astrophys. J.* **192**, L3-L6 (1974).
18. Abramowicz, M. A. & Nobili, L. *Nature* **300**, 506-508 (1982).

A fast pulsar in radio nebula CTB80

Shrinivas R. Kulkarni*, Trevor C. Clifton†, Donald C. Backer†, Roger S. Foster†, Andrew S. Fruchter‡ & Joseph H. Taylor‡

* Palomar Observatory and Owens Valley Radio Observatory, 105-24, California Institute of Technology, California 91125, USA

† Astronomy Department and Radio Astronomy Laboratory, University of California, Berkeley, California 94720, USA

‡ Joseph Henry Laboratories and Physics Department, Princeton University, Princeton, New Jersey 08544, USA

A fast pulsar with period of 39.5 milliseconds has been discovered in the peculiar combination nebula CTB80. This nebula has been proposed as the remnant of a possible historical supernova in AD 1408 (refs 1-3), but we argue that neither the nebula nor the new pulsar is linked to this supernova, which is in any case of doubtful veracity⁴. We suggest instead that the pulsar is at least 10,000 years old and that the observed features in the nebula are the result of the pulsar ploughing through the medium at $\sim 300 \text{ km s}^{-1}$. In contrast to the other young pulsars, this pulsar has a smaller magnetic field strength and is consequently under-luminous. We argue that this pulsar is more typical of young pulsars in general, a hypothesis that raises fundamental questions about the birth and evolution of pulsars.

The pulsar in CTB80 was discovered independently in the analysis of observations recorded in late June and early July by the Princeton group at Arecibo and the Berkeley/Caltech group at Green Bank⁵. Both sets of observations were motivated by the arguments of Strom⁶ that a pulsar must be present in CTB 80. The Green Bank searches were carried out on the National Radio Astronomy Observatory (NRAO) 92-m telescope, using a signal processor, developed for fast pulsar searches, consisting of a digital correlator followed by a pipeline of microprocessors

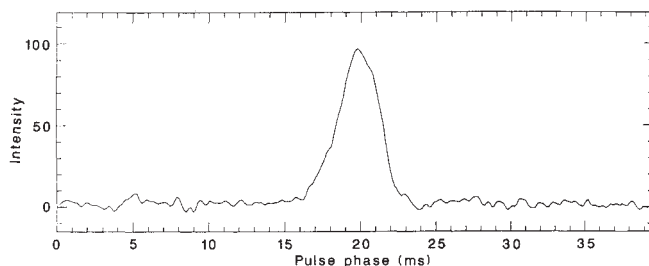


Fig. 1 Average pulse profile of PSR1951+32 from timing observation data near 1,400 MHz (ref. 7). The total bandwidth was 40 MHz and the total observing time was 172 min. Resolution of the individual profiles contributing to the average is 0.5 ms.

for real-time signal analysis. Attempts to detect the CTB80 pulsar in the standard search mode were not successful owing to insufficient integration time. A new mode was used in which 32-lag autocorrelation functions for each of two orthogonal polarizations were sampled at intervals of 0.8 ms. More than one million samples were recorded at 390 MHz with observing bandwidths of 5, 10 and 20 MHz. The data were processed by Fourier transforming the autocorrelation functions to produce 32-channel spectra, then de-dispersing the data at trial dispersion measures, and finally Fourier transforming the dedispersed data to search for periodicities. PSR1951+32 was detected in the 5-MHz data with an apparent period of 39.5288 ms at dispersion measures near 50 pc cm^{-3} .

The Arecibo observations used the National Astronomy and Ionosphere Center's (NAIC) 305-m telescope at a frequency of 1,408 MHz. The observatory's digital correlator was used as a spectrometer with fast time resolution. Four separate observations were made, each consisting of 131,072 time samples for 128 frequency channels covering a total bandwidth of 40 MHz. The first two scans used a sample integration time of 0.5 ms, while the third and fourth used integrations of 1 and 2 ms. Total

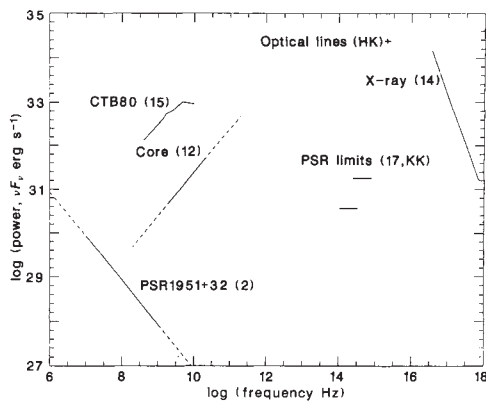


Fig. 2 Spectrum of energy output of different components of CTB80. The labelled components are discussed further in the text. Numbers in parentheses give references from text.

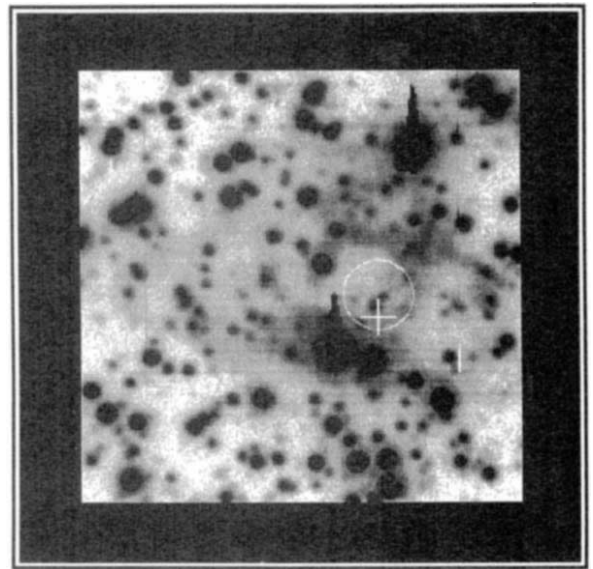


Fig. 3 An arc min square CCD (charge coupled device) image in the Gunn-r filter of the field containing the core of CTB80. The image was obtained from Palomar Observatory's Hale 5-m telescope using the four-shooter camera. North is up and east to the left. The cross marks the centre of the radio position and is based on our own optical astrometry with respect to SAO stars. The combined optical and radio uncertainty is $\sigma = 1''$ and each leg of the cross represents 3σ . The circle has a radius of $5.3''$ and represents the 99% error circle of the Einstein reprocessed position of the point X-ray source¹⁸. The bright star near the northern tip of the radio cross is candidate no. 1 of Blair and Schild¹⁸. The feature at the edge of the eastern tip is a gaseous blob. Diffuse H α of the core region is clearly visible.

observing times for the four scans thus ranged from ~ 1 –4 min. The data were processed using algorithms similar to those described above; the data were dedispersed at 512 trial dispersion measures and then searched for periodicities. PSR1951+32 was visible in all four scans with high signal-to-noise ratios. The search program yielded an apparent period, $P = 39.529 \pm 0.001$ ms and dispersion measure $DM = 50 \pm 8$ cm $^{-3}$ pc.

The discovery was followed by further observations of the pulsar at Arecibo, described in a companion paper⁷. These observations yielded a period derivative of $\dot{P} = (5.92 \pm 0.06) \times 10^{-15}$ s s $^{-1}$ and scintillation parameters that indicate a transverse velocity ~ 300 km s $^{-1}$. A pulse profile obtained from 1,400 MHz observations at Arecibo is shown in Fig. 1. The profile at 430 MHz is similar, showing a single featureless, but slightly asymmetric, peak 3 ms wide. The somewhat large duty cycle of 8% is consistent with the expectation of a $P^{-0.33}$ dependence of pulse widths in polar-cap models⁸. The absence of an interpulse in this fast pulsar is not too unusual: of the ten pulsars with $P < 100$ ms, including PSR1951+32, five have interpulses.

Application of the standard electron density model⁹ to the measured DM suggests a distance of 1.4 kpc, consistent with the lower limit derived from recent optical studies¹⁰. An earlier distance estimate of 3 kpc for the nebula¹¹ may be an overestimate. We adopt a distance of 2 kpc as a compromise between the optical and the dispersion measure estimates. The average electron density along this line of sight is then 0.023 cm $^{-3}$, which is comparable to that derived for other pulsars at similar galactic longitudes¹².

At centimetre wavelengths, CTB80 consists of three steep-spectrum ridges of length $\sim 30'$ (18 pc), at the intersection of which lies a $10' \times 6'$ (6 pc $\times$ 4 pc) plateau with a flatter spectral index of -0.4 and a roughly east-west orientation¹³. A compact $\sim 1'$ (0.6 pc) core with a very flat spectrum lies at the extreme western edge of the plateau. The continuity in the spectral index and polarization properties across these components suggests that they are physically related¹³. CTB80 is classified as a combination¹⁴ or composite¹⁵ supernova remnant (SNR) because the core is reminiscent of the pulsar-driven Crab nebula, and the ridges, though highly irregular, can be considered parts of a supernova shell remnant.

The core consists of a polarized, limb-brightened, incomplete shell, immersed in diffuse emission elongated in the E-W direction^{1,16}. In Strom's 20-cm VLA (Very Large Array) map⁶, PSR1951+32 is seen as a highly polarized, steep spectrum ($\alpha \sim -2$) point source close to the western limb of this shell. A bright hot spot is found in the limb southwest of the pulsar. Strong optical emission (H α , [O III], and [S II]) is also found in the core region with similar morphology^{10,11}. X-ray observa-

tions made with the Einstein Observatory show both diffuse X-ray emission, roughly coincident with the plateau, and a point source with strength comparable to the diffuse emission². The X-ray point source coincides, within astrometric uncertainties, with the pulsar.

The discovery of PSR1951+32 convincingly identifies the source of energy injection in the core as the rotational energy loss of the pulsar: $\dot{E}_r = I\Omega\dot{\Omega} \approx 3.7 \times 10^{36}$ erg s $^{-1}$, where $\Omega = 2\pi/P$. Assuming a flat radio spectrum up to 20 GHz, the core radio luminosity is 5×10^{31} erg s $^{-1}$, based on the observed flux density of 500 mJy at centimetre wavelengths¹. The radio luminosity of the entire CTB80 radio nebula is estimated as 10^{33} erg s $^{-1}$ (ref. 17). The total luminosity from the bright filaments and the diffuse gas is $\leq 2 \times 10^{34}$ erg s $^{-1}$ (J. Hester and S.R.K. (HK in Fig. 3), in preparation). The combined luminosity of the X-ray point source and diffuse emission is 1.4×10^{34} erg s $^{-1}$ (ref. 2). The power spectra of the components of CTB80 are summarized in Fig. 2. The pulsar is thus clearly capable of supplying enough energy to account for these observed energy losses.

The power needed to account for the observed pulsed emission is insignificant by comparison. In the radio window, assumptions of a low frequency cutoff at 10 MHz and isotropic emission lead to a luminosity of 8×10^{29} erg s $^{-1}$. A faint ($m_v \approx 20$) optical star, which has been considered a counterpart to this X-ray source¹⁸ (see Fig. 3), is most likely a foreground star owing to its magnitude and limits on optical pulsations (J. Middleditch, personal communication; J. Kristian and S.R.K. (KK in Fig. 3), in preparation). Upper limits to pulsed fluxes at infrared¹⁹ and optical wavelengths at the position of the pulsar imply luminosities no greater than 4×10^{30} erg s $^{-1}$ and 2×10^{31} erg s $^{-1}$, respectively in these bands.

The origin of the optical lines in the nebular emission is uncertain. Both shocks and photoionization models have been

considered¹⁰. This uncertainty is unfortunate, because a proper discussion of the energy budget must include allowance for energy losses through all channels including losses through the as yet unobserved collisionally excited infrared and near-ultraviolet lines. In the photoionization model, every H α photon implies approximately two recombinations²⁰ and hence, in a steady state, two ionizations. The observed H α luminosity¹⁰ of 2.7×10^{33} erg s⁻¹ then implies a power of at least 4×10^{34} erg s⁻¹, and more likely $\sim 10^{35}$ erg s⁻¹, in ionizing soft X-rays.

While the implied flux in soft X-rays is only $0.03 \dot{E}_r$, this assumed synchrotron spectrum must turnover between 200 eV and 2 eV in order to satisfy the absence of an optical synchrotron nebula. For a range of spectral indices consistent with the Einstein observations ($-2 > \alpha > -4$, ref. 2) and a range of turnover frequencies, we predict an optical synchrotron nebula with an integrated magnitude (including two magnitudes of extinction) of 14 mag. The photoionization model can be tested by future deep optical searches.

We now discuss the age of the pulsar. Assuming a constant magnetic moment, the true age, t , of a pulsar dominated by magnetic dipole radiation losses is $t = \tau_c [1 - (P_0/P)^2]$ where P_0 is the period at birth, P is the current period and $\tau_c = P/2\dot{P}$ is the so-called characteristic age. Thus τ_c is an upper limit to the age of the pulsar. Strom and others¹⁻³ have argued that CTB 80 is connected with a possible historical supernova of AD 1408. If this is true, PSR1951+32 was born 579 yr ago with a period approximately 99.7% of the current period. The point X-ray source could then be explained as thermal emission from the surface of the neutron star at a temperature $T \approx 10^6$ K. We question this hypothesis for two reasons. First, the predicted velocities of the nebular material are an order of magnitude larger than the observed velocities¹⁰ and the velocities needed to explain the ridge are relativistic²¹. Second, even the historical data have been questioned⁴. We see no compelling reason to link PSR1951+32 to a possible historical supernova in AD 1408.

We suggest instead that PSR1951+32 is at least 10^4 yr old, and that the plateau region is a wake of either shocked interstellar medium (ISM) or of relativistic particles and magnetic field which were left by the pulsar as it ploughed through the ISM with a speed $v \approx 300$ km s⁻¹. The hot spot southwest of the pulsar⁶ locates the region where the pulsar wind is stopped and randomized. The $>10'$ extent of the plateau, together with the adopted distance of 2 kpc and the measured scintillation speed⁷, yield a minimum pulsar age of $\geq 10^4$ yr. The observed increase in spectral index with increasing distance from the core^{2,13} can be explained by the pulsar wake model as a consequence either of cooling or radiative and adiabatic expansion losses.

A 'pulsar-wind nebula' model^{22,23}, which was developed to explain the X-ray synchrotron nebulae surrounding old, radio pulsars with large rotational energy loss rates, can be applied to the pulsar-core region of CTB80 (see also ref. 24). Owing to the large space motions the magnetized plasma wind from a pulsar could be confined by the ram pressure of the intersellar medium. Synchrotron radiation from the confined plasma gives rise to the diffuse X-rays. When the ram pressure is unable to confine the pulsar wind, adiabatic losses dominate and the nebula disappears. In either case, a compact source, which results from synchrotron emission of the wind, is expected at the location of the pulsar²². The pressure in the core region, from minimum energy arguments, is estimated⁶ to be $\sim 3 \times 10^{-9}$ erg cm⁻³. Thus a mean ambient ISM density, $n_H \geq 4(v/300 \text{ km s}^{-1})^{-2} \text{ atom cm}^{-3}$, is sufficient to confine the wind from PSR1951+32. The rough alignment of the X-ray nebula with the radio plateau region, as well as the location of the radio hot spot, suggest that the pulsar is moving from east to west⁶. If the pulsar does not exhibit glitches or significant timing noise, timing observations over the next few years should be able to determine the magnitude and orientation of the proper motion.

As an aside, we note that the poor quality of the Einstein data does not allow firm proof that the X-ray emission is non-

thermal on any scale. Indeed, Strom *et al.*¹ invoke a thermal origin for the X-rays in order to explain the observed depolarization at radio wavelengths. However, recently discovered faint, spatially extended ionized gas (Hester & Kulkarni, in preparation) will certainly provide the necessary Faraday depth for depolarization, provided it is mixed with the synchrotron particles. Hence we believe there is no compelling reason to invoke a thermal origin for the X-ray emission.

At present, PSR1951+32 appears to be in a high mean density region as evidenced by copious optical line emission. The absence of any anomalous abundances or high velocities leads us to conclude that the ambient gas is not circumstellar gas left over from the supernova explosion that formed the pulsar. If PSR1951+32 is indeed only 10^4 yr old it is hard to understand the presence of this rich ISM because the explosion would have evacuated a cavity. The simplest conclusion is that PSR1951+32 is considerably older than 10^4 yr, but of course less than $\tau_c \approx 10^5$ yr. Then the pulsar can overtake the expanding supernova shell and move out into the general ISM. In this scenario, the pulsar was born $\geq 1^\circ$ east of its current position, and the eastern edge of the plateau marks the onset of high density interstellar gas. It is tempting to associate the ridges with the remains of the supernova explosion at the birth of PSR1951+32. The lack of a reasonably symmetric shell can be attributed to the distribution of dense molecular clouds. Indeed, there appears to be some ¹²CO emission in this area in the low resolution maps of Dame *et al.*²⁵. Clearly a detailed study of the ISM in this region of sky is highly desirable.

The conventional model for the origin and the evolution of neutron stars is based on the idea that they are born in type II supernovae with short periods ($P \leq 20$ ms) and strong magnetic fields ($B \geq 10^{12}$ G) that decay on timescales of several million years. Almost all of these statements have been challenged. First, of the ~ 75 type II galactic SNRs, only five (including CTB80) have been clearly associated with a neutron star²⁶. Second, statistical analyses of the pulsar population²⁷ and of pulsar surveys²⁸⁻³⁰ suggest that many pulsars are 'injected' into the population with large periods ($P > 500$ ms) and high magnetic field strengths (but other analyses do not find any need for 'injection')³¹. Finally, while all the analyses agree that the distribution of magnetic field of very young pulsars is tight, there is no consensus on the median field value. The median field value ranges from 0.7×10^{12} G (ref. 9) to 3×10^{12} (refs 29-31) and appears to depend upon the details of the assumed evolution of luminosity. Thus in the models of Narayan³⁰ and Stollman³¹, PSR1951+32, a young pulsar with a field strength of only 0.5×10^{12} G, is unusual, whereas in the Lyne *et al.*⁹ model, it is a typical young pulsar.

Regardless of this uncertainty, the discovery of PSR1951+32 shows that not all pulsars are born with strong magnetic fields like the Crab pulsar. This finding has important repercussions because low magnetic field pulsars are expected to be less luminous than high magnetic field pulsars, both at X-ray and radio wavelengths. Clearly, it is important to determine observationally the fraction of pulsars that are born like PSR1951+32.

We first consider the X-ray data. The combined X-ray luminosity of the point source and the nebula around PSR1951+32 is comparable to that of other unidentified X-ray sources thought to be associated with supernova remnants^{15,26}. This fact, coupled with the absence of sources as luminous as the Crab pulsar and nebula, suggests that the CTB80 X-ray nebula plus PSR1951+32 may be a better representative of young active remnants than the Crab. The low luminosity of young active SNRs could be due either to pulsars with short P and low B or to longer P , high B pulsars. Indeed, both possibilities have now been observed: PSR1951+32 in CTB80 and PSR1508-59 in MSH15-52 respectively.

Radio data show that the mean luminosity of PSR1951+32 at 400 MHz is 40 mJy kpc², much fainter than the Crab and Vela pulsars, with luminosities of 3,000 and 1,000 mJy kpc²,

respectively. Most existing pulsar surveys have discriminated against short period pulsars, and all surveys are sensitivity limited. Together with a pulsar beaming factor less than unity, these facts are consistent with the suggestion that many active SNRs may contain pulsars like PSR1951+32. Indeed, Helfand and Becker¹⁵ argue on the basis of total energy content that all active remnants must contain a fast pulsar (initial period ≈ 20 ms) and that varying initial conditions, such as magnetic field and density of the local interstellar medium, are responsible for the wide range of properties. As emphasized by Lyne *et al.*⁹, sensitivity limits ensure that low-luminosity pulsars are heavily under-represented in the sample of known pulsars, relative to the galactic population as a whole. Thus, it is possible that among the overall population PSR1951+32 may be much more typical of young pulsars than the Crab and Vela pulsars.

This conclusion raises some interesting issues because most observed pulsars, despite being considerably older than PSR1951+32, have magnetic field strengths several times larger. If the field of PSR1951+32 is typical of young pulsars, it provides evidence that pulsar magnetic fields can grow with time. Theoretical scenarios of field amplification by a thermoelectric instability have been proposed³². The existing model is probably incomplete because, in contrast to the predicted $\tau_m \approx 10^5$ yr growth timescale, there are four young neutron stars (Crab, Vela, CTB109, MSH15-52) already known to possess strong magnetic fields. In passing, we note that the assumption of an exponential field growth increases the upper limit for the true age of PSR1951+32 by at most a factor of two, and thus has little impact on the discussion regarding the age of the nebula.

The discovery of PSR1951+32 clearly demonstrates that the parameters of pulsars at birth can vary widely, and should encourage further work on the distribution of initial pulsar periods and magnetic field strengths. This is probably best done by a combination of X-ray imaging and further sensitive pulsar searches.

We thank P. Perillat, M. Ryba and D. Stinebring for help with the Arecibo observations, and A. Dey and L. Fairhead for help with Green Bank observations. We also thank L. Anacker, E. Hook and B. Siebers for software development assistance at the John von Neumann Center at Princeton. We gratefully acknowledge useful discussion with J. Arons, J. Hester S. Phinney and R. Romani. The Arecibo Observatory is part of NAIC, which is operated by Cornell University under contract with the NSF. NRAO is operated by Associated Universities Inc. under contract with the NSF. ASF is supported by a von Neumann fellowship at Princeton University.

Received 27 August; accepted 21 October 1987.

1. Strom, R. G., Angerhofer, P. E. & Dickel, J. R. *Astr. Astrophys.* **139**, 43-49 (1984).
2. Wang, Z. R. & Seward, F. D. *Astrophys. J.* **285**, 607-612 (1984).
3. Strom, R. G., Angerhofer, P. E. & Velusamy, T. *Nature* **284**, 38-40 (1980).
4. Stephenson, F. R. & Yau, K. K. C. *Q. J. R. astr. Soc.* **27**, 559-568 (1986).
5. Clifton, T. R. *et al.* *IAU Circ. No.* 4422 (1987).
6. Strom, R. G. *Astrophys. J.* **319**, L103-L108 (1987).
7. Fruchter, A. S. *et al.* *Nature* **330**, 53-54 (1987).
8. Ruderman, M. A. & Sutherland, P. G. *Astrophys. J.* **196**, 51-72 (1975).
9. Lyne, A. G., Manchester, R. N. & Taylor, J. H. *Mon. not. R. astr. Soc.* **213**, 613-639 (1983).
10. Blair, W. P., Kirshner, R. P., Fesen, R. A. & Gull, T. R. *Astrophys. J.* **282**, 161-171 (1984).
11. Angerhofer, P. E., Wilson, A. S. & Mould, J. R. *Astrophys. J.* **236**, 143-152 (1980).
12. Weisberg, J. M., Rankin, J. & Boriakoff, V. *Astr. Astrophys.* **88**, 84-93 (1980).
13. Angerhofer, P. E., Strom, R. G., Velusamy, T. & Kundu, M. R. *Astr. Astrophys.* **94**, 313-322 (1981).
14. Weiler, K. W. *IAU Symp.* **101**, 299-320 (1983).
15. Helfand, D. J. & Becker, R. H. *Astrophys. J.* **314**, 203-214 (1987).
16. Velusamy, T. & Kundu, M. R. *J. Astr. Astrophys.* **4**, 253-260 (1983).
17. Mantovani, F., Reich, W., Salter, C. J. & Tomasi, P. *Astr. Astrophys.* **145**, 50-58 (1985).
18. Blair, W. P. & Schild, R. E. *Astrophys. Lett.* **24**, 189-195 (1985).
19. Clifton, T. R. *et al.* preprint (1987).
20. Osterbrock, D. E. *Astrophysics of Gaseous Nebulae* (Freeman, San Francisco, 1974).
21. van den Bergh, S. *Publs astr. Soc. Pacif.* **92**, 768-770 (1980).
22. Cheng, A. F. *Astrophys. J.* **275**, 790-801 (1983).
23. Cheng, A. F. in *Workshop on Millisecond Pulsars* (eds Reynolds, S. P. & Stinebring, D. R.) 294-304 (NRAO, Green Bank, 1984).
24. Arons, J. *Nature* **302**, 301-305 (1983).
25. Dame, T. M. *et al.* *Astrophys. J.* (in press).
26. Helfand, D. J. & Becker, R. H. *Nature* **307**, 215-221 (1984).
27. Vivekanand, M. & Narayan, R. *J. Astr. Astrophys.* **2**, 315-337 (1981).

28. Chevalier, R. A. & Emmering, R. T. *Astrophys. J.* **304**, 140-153 (1986).
29. Narayan, R. *Astrophys. J.* **319**, 162-179 (1987).
30. Narayan, R. & Schaudt, K. J. *Astrophys. J.* (submitted).
31. Stollman, G. M. *Astr. Astrophys.* **178**, 143-152 (1987).
32. Blandford, R. D., Applegate, J. H. & Hernquist, L. *Mon. Not. R. astr. Soc.* **204**, 1025-1048 (1983).

Timing and scintillation observations of the fast pulsar in CTB80

A. S. Fruchter*, J. H. Taylor*, D. C. Backer†, T. R. Clifton†, R. S. Foster† & A. Wolszczan‡

*Joseph Henry Laboratories and Physics Department, Princeton University, Princeton, New Jersey 08544, USA

†Astronomy Department, University of California, Berkeley, California 94720, USA

‡Arecibo Observatory, Arecibo, Puerto Rico 00612, USA

Shortly after the discovery¹ of a 39.5-ms pulsar in the supernova remnant CTB 80, we began a series of timing measurements at the Arecibo Observatory to obtain more accurate values for the period and dispersion measure and to determine the period derivative and scintillation parameters. Data acquired over a span of 11 days yielded a surprisingly small period derivative, $\dot{P} = (5.92 \pm 0.06) \times 10^{-15} \text{ s s}^{-1}$. If the pulsar slows by magnetic dipole radiation, this spin-down rate implies a surface field strength of only 5×10^{11} gauss yet the pulsar is believed to be much too young for its magnetic field to have decayed. Scintillation observations reveal unusually rapid fluctuations, indicating a large pulsar transverse velocity. Here we describe the timing and scintillation measurements and their results, and briefly explore some astrophysical consequences.

Timing measurements of PSR1951+32, the pulsar in CTB 80, were made with the Arecibo 305-m telescope on the nights of 17, 20, 22 and 26-28 July at observing frequencies near 1,408 MHz. Signals of both circular polarizations were amplified by cooled GaAs FET pre-amplifiers, converted to an intermediate frequency, and detected in a $2 \times 32 \times 0.25$ MHz filter bank spectrometer. The output of each filter channel, summed over polarizations, was sampled at 512 times the apparent pulsar rotation frequency by one of 32 synchronously operating signal averagers^{2,3}. In this way the pulsar's periodic waveform was accumulated into 32 integrated pulse profiles, each with 512 equally spaced phase bins. The profiles were recorded in digital form every two minutes, together with a fiducial time recorded from the observatory's master clock at a specified phase in the signal averager cycle.

The profiles were later corrected for differential dispersion delays (which amount to 1.1 ms across the 8 MHz total bandwidth), summed over the 32 channels, and reduced to effective pulse arrival times by template matching in the Fourier-transform domain⁴. Typical uncertainties for arrival times based on the two-minute integrations were $\sim 275 \mu\text{s}$. A total of 229 such measurements were obtained on the six observing nights.

Topocentric pulse arrival times were transformed to the reference frame of the Solar System barycentre, and the resulting times⁵ used to determine the pulsar period, P , derivative, $\dot{P}$, and dispersion measure, DM, by the method of least squares. We obtained the results for Julian date 2446993.66

$$P = 0.03952975206 \pm 3 \text{ s}$$

$$\dot{P} = (5.92 \pm 0.06) \times 10^{-15} \text{ s s}^{-1}$$

$$\text{DM} = 45 \pm 2 \text{ cm}^{-3} \text{ pc}$$

(The uncertainty quoted for the period refers to the last digit.) For barycentric reductions we used the position quoted for the pulsar candidate by Strom⁶, namely $\alpha(1,950) =$

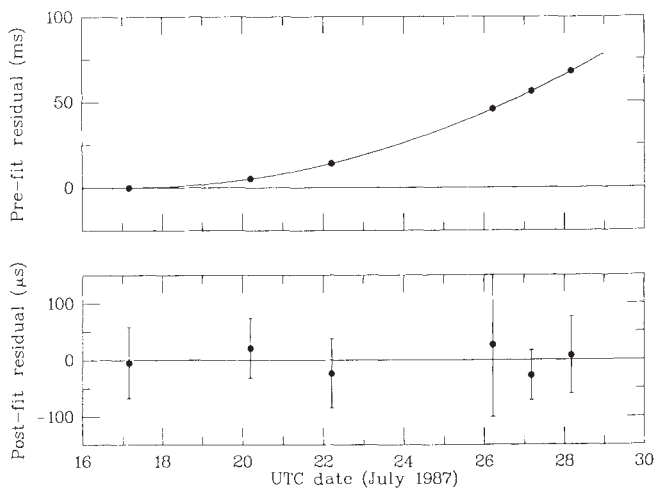


Fig. 1 Pre- and post-fit residuals for the timing solution based on a single phase measurement for each of six observing nights. At the top, pre-fit residuals (computed assuming $\dot{P} = 0$) illustrate the fact that the pulsar is slowing down; the parabola drawn through the points corresponds to the solution quoted for P and $\dot{P}$. The lower panel shows the post-fit residuals on an expanded scale.

19 h 51 min 2.47 s, $\delta(1950) = +32^\circ 44' 50.2''$. Uncertainties in this position, about $1''$ in each coordinate, contribute negligibly to errors in our fitted parameters.

Figure 1 illustrates the pre- and post-fit residuals from the timing solution, with the data averaged to yield a single effective arrival time for each night's observations. The pre-fit residuals, shown in the upper portion of the diagram, were computed under the assumption $\dot{P} = 0$. After an interval of 11 days, pulses from PSR1951+32 were arriving nearly two periods later than would be expected if the period were constant. Post-fit residuals, plotted in the lower panel of Fig. 1 on a scale expanded by a factor of 300, are consistent with random measurement errors of $\sim 50 \mu\text{s}$ in each night's measured pulsar phase.

Additional measurements of the pulse arrival times were made using the Arecibo autocorrelation spectrometer in a recently developed signal-averaging mode (T.R.C., D. A. Frail, S. R. Kulkarni & J. M. Weisberg, in preparation) which permits the rapid accumulation of radio spectra. The spectra were synchronously averaged at the pulsar period to yield a two-dimensional array representing signal intensity as a function of both frequency and pulsar phase. Sixty four spectral channels were used, covering a total bandwidth of 40 MHz at 1,400 MHz and 5 MHz at 430 MHz. We sampled the spectra at 0.5 ms intervals and time-tagged the data with a signal from the Observatory master clock. These observations provided the pulsar profile shown in the accompanying article¹, as well as pulse arrival times consistent with the parameters quoted above.

In order to obtain an accurate estimate of the scintillation bandwidth and timescale, a dynamic spectrum of the pulsar was measured on 5 August at 1,400 MHz using the autocorrelator in the mode described above (for further details of the scintillation observing procedure and analysis see ref. 7). We sampled 256 spectral channels across the 40 MHz bandpass, integrating

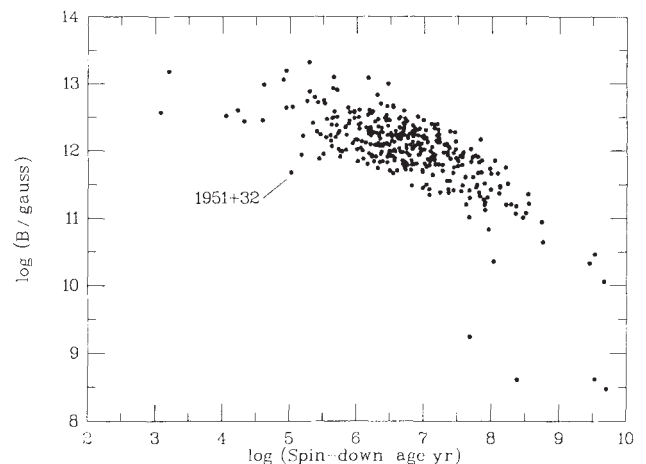


Fig. 2 Surface magnetic field strengths plotted as a function of spin-down age for 366 pulsars. PSR1951+32 has the weakest field among young pulsars, and we suggest that its field may still be in the process of turning on.

each of the 128 consecutive spectra for 20 s. Two-dimensional autocorrelation analysis of the resultant dynamic spectrum revealed a characteristic decorrelation bandwidth (half width at half maximum of the autocorrelation function) of 1.1 ± 0.3 MHz, and a characteristic timescale of 38 ± 3 s. These parameters, taken together with a distance to CTB 80 of ~ 2 kpc, a value consistent with both the pulsar dispersion measure and estimates¹ of the distance to the remnant, imply⁸ a pulsar transverse velocity of approximately 300 km s^{-1} . While a scintillation-derived velocity is model dependent, our measurement suggests that the pulsar's speed is unusually high. Evidence that the large apparent velocity is not caused by anomalous scattering comes from the work of Wolszczan and Sieber (in preparation) who find, for pulsars with dispersion measures near $45 \text{ cm}^{-3} \text{ pc}$, typical 1,400 MHz decorrelation bandwidths of 1–4 MHz and decorrelation timescales of 100–600 s.

The apparent dipolar magnetic field at the neutron star surface, 5×10^{11} gauss (estimated⁵ from the relation $B = (\dot{P}\dot{P}_{-15})^{1/2} \times 10^{12}$ gauss, where $\dot{P}_{-15} = 10^{15} \dot{P}$), is less than that of 306 of the 366 pulsars for which $\dot{P}$ is known, and is the smallest among all pulsars with observed spin-down ages $\dot{P}/2\dot{P} < 10^6$ yr (see Fig. 2). However, PSR1951+32 is probably no older than its spin-down age, 1.1×10^5 yr, and is perhaps as young⁶ as 580 yr. It is certainly far too young for its magnetic field to have decayed significantly⁹. An interesting possibility is that, far from decaying, this pulsar's magnetic field may still be turning on by thermal generation¹⁰. If this is true, the period second derivative will be positive and relatively large, and should be measurable within about two years. Over the course of time the pulsar would then evolve upward and to the right in Fig. 2, perhaps joining the main group of points in the diagram is less than a million years.

We thank the staff of the Arecibo Observatory, especially M. Davis, G. Giles and P. Perillat, for assistance with the observations and with rapid transport of the data. The Arecibo Observatory is part of the National Astronomy and Ionosphere Center, which is operated by Cornell University under contract with the NSF.

Received 21 August; accepted 1 October 1987.

1. Kulkarni, S. R., Clifton, T. R., Backer, D. C., Foster, R. S., Fruchter, A. S. & Taylor, J. H. *Nature* **330**, 50–53 (1987).
2. Rawley, L. A., Taylor, J. H., Davis, M. M. & Allen, D. W. *Science* (in the press).
3. Rawley, L. A., Taylor, J. H. & Davis, M. M. *Astrophys. J.* (in the press).
4. Rawley, L. A. thesis, Princeton University (1986).

5. Manchester, R. N. & Taylor, J. H. *Pulsars* (Freeman, San Francisco, 1977).

6. Strom, R. G. *Astrophys. J.* (in the press).

7. Wolszczan A. & Cordes J. M. *Astrophys. J.* **320**, L35–L38 (1987).

8. Cordes, J. M. *Astrophys. J.* **311**, 183–196 (1986).

9. Lyne, A. G., Manchester, R. N. & Taylor, J. H. *Mon. Not. R. astr. Soc.* **213**, 613–639 (1985).

10. Blandford, R. D., Applegate, J. H. & Hernquist, L. *Mon. Not. R. astr. Soc.* **204**, 1025–1048 (1983).

Vertical oceanic nutrient fractionation and glacial/interglacial CO₂ cycles

Edward A. Boyle

Department of Earth, Atmospheric and Planetary Sciences,
Room E34-258, Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, USA

Ice-core studies have established that atmospheric carbon dioxide is about 80 parts per million by volume (p.p.m.v.) lower during cold glacial climates than it is during warm interglacial times¹⁻⁸. A number of models have been offered for this phenomenon⁹⁻¹². These models are consistent with observation of an enhanced carbon isotope contrast between surface and deep waters, but they also predict dissolved oxygen depletions which are inconsistent with the widespread occurrence of glacial-age fossils from aerobic benthic organisms. Recent observations on changes in the vertical oceanic chemical structure provide a new mechanism to control glacial/interglacial CO₂ change. Palaeochemical evidence shows that nutrients and metabolic CO₂ are shifted from intermediate waters into deeper waters during glacial times¹³⁻¹⁵. At the outset of this redistribution, increased deep dissolved CO₂ concentrations raise carbonate dissolution above that required for a steady-state input/output balance. Oceanic alkalinity then increases until steady-state dissolution rates are restored. Here I propose that atmospheric CO₂ decreases in direct response to increased oceanic alkalinity. Although the carbonate response factor in atmospheric CO₂ has been noted before, it has not been linked with vertical nutrient rearrangements. The withdrawal of nutrients from intermediate waters of the ocean also prevents oxygen depletion in these waters.

Data on the nutrient-related properties $\delta^{13}\text{C}$ and Cd in benthic foraminifera from cores representing the depth range 1,700–4,200 m in the western North Atlantic show that the intermediate-depth glacial North Atlantic was nutrient-depleted relative to today, and deeper waters were more nutrient enriched (refs 13–15 and L. Peterson, personal communication). The Atlantic intermediate-depth nutrient depletion is equivalent to a decrease of $\sim 0.6 \mu\text{mol kg}^{-1}$ in phosphorus. The situation is not as clear for the Pacific (because of a paucity of suitable sediment cores), but J. C. Duplessy *et al.* (manuscript in preparation) have obtained data showing that intermediate waters of the North Pacific above 2,500 m also became more nutrient-depleted during glaciation. The Pacific glacial/interglacial carbon isotope shift is about +0.4‰, equivalent to about $-0.5 \mu\text{mol kg}^{-1}$ in phosphorus. So it appears that in both the Atlantic and the Pacific, nutrients are transferred from intermediate waters into deeper waters during glacial times.

As outlined above, a decrease in atmospheric CO₂ results from the effect of increased deep CO₂ on carbonate dissolution. The excess CO₂ initially increases the extent of carbonate dissolution in the deep ocean to a level above that required to maintain the steady-state carbonate input/output balance. Two simplifying assumptions allow for direct calculation of the magnitude of glacial/interglacial atmospheric CO₂ changes: (1) the deepest half of the ocean is considered homogeneous; and (2) all sinking particulate biogenic matter is considered to degrade in 'Redfield' proportions 106 CH₂O:21 CaCO₃:1 P (ref. 16). Then, for a given change in the deep phosphorus content (ΔP), deep total dissolved carbon dioxide ($\Delta\Sigma\text{CO}_2 \sim \Delta(\text{HCO}_3^-) + \Delta(\text{CO}_3^{2-})$) and alkalinity ($\Delta\text{Alk} \sim \Delta(\text{HCO}_3^-) + 2\Delta(\text{CO}_3^{2-})$) will initially change by:

$$\Delta\Sigma\text{CO}_2 \approx 127 \Delta\text{P} \quad (1)$$

$$\Delta\text{Alk} \approx 42 \Delta\text{P} \quad (2)$$

and the carbonate ion content will drop by the difference between the added alkalinity and dissolved carbon dioxide

content:

$$\Delta[\text{CO}_3^{2-}] \approx \Delta\text{Alk} - \Delta\Sigma\text{CO}_2 = -85 \Delta\text{P} \quad (3)$$

As Broecker^{9,17} pointed out some time ago, steady-state carbonate saturation (that is, no change in deep $[\text{CO}_3^{2-}]$) must eventually be restored to match this initial $[\text{CO}_3^{2-}]$ decrease. Mean oceanic alkalinity increases through dissolution of $85 \mu\text{mol kg}^{-1}$ CaCO₃. The new atmospheric CO₂ content resulting from this alkalinity increase can be calculated by substituting into the thermodynamic equations for p_{CO_2} of surface waters as a function of ΣCO_2 and Alk¹⁸.

The average change in intermediate-water nutrient content is about $0.6 \mu\text{mol kg}^{-1}$; as the volume of the ocean above 2,500 m is nearly equal to the volume below that depth¹⁹, deep-ocean phosphorus must have increased by the same amount. The predicted rise in oceanic alkalinity results in a 40 p.p.m.v. drop in atmospheric CO₂. Thus the magnitude of the CO₂ change calculated in this fashion accounts for about half of the polar ice-core observations. The residual decrease in atmospheric CO₂ might be due to underestimation of intermediate-water nutrient depletion, as very little glacial intermediate-water data is available as of yet. The residual also could be due to contributions from mechanisms suggested previously.

The above estimate for glacial CO₂ change is relatively model-independent, because it is determined only by the resultant chemical structure of the ocean. The estimate would be valid for any model which produces the same chemical structure. A similar estimate has been derived from a number of different multiple box models of the ocean, which will be presented elsewhere. It is likely that many models can be produced which predict the same nutrient redistribution; all would predict similar consequences for atmospheric CO₂. Numerous consequences of this process should be preserved by marine sediments; in particular, preservation histories of calcium carbonate will be affected by this process. These effects will be considered in detail elsewhere.

The carbon isotope ratio of the deep ocean drops by -0.6% at the same time as the vertical nutrient rearrangement. Hence variations in the benthic-planktonic $\Delta\delta^{13}\text{C}$ contrast^{9,20} need not be due to changes in phosphorus content of the ocean or to changes in the pre-formed phosphorus content of polar ocean waters; they may simply result from a shift of nutrients between the intermediate and deep ocean.

Although Broecker and Peng²¹ did not address the consequences of the transfer of nutrients and CO₂ from intermediate waters into the deep ocean, they did consider aspects of the interaction between carbonate saturation and atmospheric CO₂. In particular, they pointed out that there is a lag of several thousand years in the oceanic alkalinity response relative to the initial perturbation. In the present model, the initial driving force is reorganization of the vertical chemical structure of the ocean. Hence it is envisioned that changes in atmospheric CO₂ content will lag several thousand years behind the new oceanic chemical structure and, likewise, that the $\delta^{13}\text{C}$ difference between surface and deep waters ($\Delta\delta^{13}\text{C}$) will lead atmospheric CO₂ change. This lag accounts for the observed lead of $\Delta\delta^{13}\text{C}$ relative to benthic $\delta^{18}\text{O}$ (ref. 22). This lag also accommodates the observation that minimum $\Delta\delta^{13}\text{C}$ appears to have occurred about 7,000–12,000 yr BP (ref. 23 and A. Mix and N. Shackleton, personal communication), considerably before full interglacial CO₂ was attained^{2,5,24}.

The consequences of vertical oceanic chemical reorganization for oxygen can be estimated by assuming oxygen consumption of between 135 to 170 molecules for every phosphorus atom regenerated in the deep ocean¹⁶. Currently, the oxygen concentration of intermediate waters (100–2,500 m) averages $\sim 110 \mu\text{mol kg}^{-1}$; deep-ocean oxygen averages $\sim 190 \mu\text{mol kg}^{-1}$ (ref. 25). In the absence of major changes in ocean temperature, the nutrient redistribution proposed above will alter oxygen

levels by $\sim 80\text{--}100\ \mu\text{mol kg}^{-1}$. Hence, mean intermediate oxygen levels would rise to $190\text{--}210\ \mu\text{mol kg}^{-1}$ and deep ocean concentrations would fall to $90\text{--}110\ \mu\text{mol kg}^{-1}$: the oxygen crisis that occurs in previous models is averted. Only in the most oxygen-depleted deep waters of the ocean would near-anoxia occur. The oxygen content of the Panama Basin is $115\ \mu\text{mol kg}^{-1}$, so these waters would approach oxygen depletion. If the deep waters of the ocean were $\sim 2^\circ\text{C}$ cooler²⁶, then deep-water oxygen concentrations of the Panama Basin during glacial times would have been $20\text{--}40\ \mu\text{mol kg}^{-1}$. Also, as the Panama Basin is rapidly (~ 50 yr) ventilated by mid-depth waters over a relatively shallow sill ($2,920\text{ m}$)²⁷, the inflowing waters may have somewhat higher initial oxygen levels during glacial time.

While the average deep-ocean basin is not in danger of oxygen depletion during glacial times, this model still predicts that deep-ocean oxygen decreases significantly. Any sedimentary process which depends on the bottom-water oxygen content should record these predicted changes. Using the model developed by Emerson²⁸, the predicted decrease in deep-water oxygen could account for the increased organic carbon content of Panama Basin sediments²⁹. The predicted decrease in oxygen with depth might also account for the increase in organic carbon content with depth observed in glacial age sediments from the eastern equatorial Atlantic³⁰. Another consequence of this model is that the oxygen content of the mid-depth ocean should have increased, so we would expect to find lower organic carbon contents in glacial-age sediments from intermediate-depth ocean cores (in regions where surface productivity did not increase). These oxygen concentration changes also have implications for the sedimentary cycle of redox-sensitive elements such as manganese.

In conclusion, it is clear that the vertical chemical structure

of the ocean between glacial and interglacial times is a significant control on late Quaternary atmospheric CO_2 variability. Future observational work should endeavour to clarify the extent of this vertical restructuring and to examine the sedimentary record for the traces of this process, in particular those relating to the predicted lag of the CO_2 response relative to the oceanic forcing. I thank Laurent Labeyrie and J. R. Toggweiler for stimulating discussions. This report grew out of work sponsored by the NSF.

Received 28 September; accepted 20 November 1987.

1. Barnola, J. M., Raynaud, D., Neftel, A. & Oeschger, H. *Nature* **303**, 410–413 (1983).
2. Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
3. Berner, B., Stauffer, B. & Oeschger, H. *Nature* **275**, 53–55 (1979).
4. Delmas, R. J., Ascencio, J.-M. & Legrand, M. *Nature* **284**, 155–157 (1980).
5. Neftel, A., Moor, E., Oeschger, H. & Stauffer, B. *Nature* **315**, 45–47 (1985).
6. Neftel, A., Oeschger, H., Schwander, J., Stauffer, B. & Zumbund, R. *Nature* **295**, 220–223 (1982).
7. Oeschger, H. *et al.* *Am. geophys. Un. Monogr.* **29**, 299–306 (1984).
8. Oeschger, H., Stauffer, B., Finkel, R. & Langway, C. C. *Am. geophys. Un. Monogr.* **32**, 132–142 (1985).
9. Broecker, W. S. *Geochim. cosmochim. Acta* **46**, 1689–1705 (1982).
10. Sarmiento, J. L. & Toggweiler, J. R. *Nature* **308**, 621–624 (1984).
11. Siegenthaler, U. & Wenk, Th. *Nature* **308**, 624–625 (1984).
12. Knox, S. & McElroy, M. J. *geophys. Res.* **89**, 4629–4637 (1984).
13. Boyle, E. & Keigwin, L. D. *Nature* **330**, 35–40 (1987).
14. Zahn, R., Sarnthein, M. & Erlenkeuser, H. *Paleoceanography* (in the press).
15. Oppo, D. & Fairbanks, R. G. *Earth planet. Sci. Lett.* (in the press).
16. Takahashi, T., Broecker, W. S. & Langer, S. J. *geophys. Res.* **90**, 6907–6924 (1985).
17. Broecker, W. S. *Prog. Oceanogr.* **11**, 151–197 (1982).
18. Edmond, J. M. & Gieskes, J. M. *Geochim. cosmochim. Acta* **34**, 1261–1291 (1970).
19. Menard, H. W. & Smith, S. M. J. *geophys. Res.* **71**, 4305–4326 (1966).
20. Shackleton, N. J., Hall, M. A., Line, J. & Shuxi, C. *Nature* **306**, 319–322 (1983).
21. Broecker, W. S. & Peng, T. H. *Global biogeochem. Cycles* **1**, 15–30 (1987).
22. Shackleton, N. J. & Pisias, N. G. *Am. geophys. Un. Monogr.* **32**, 313–318 (1984).
23. Curry, W. B. & Crowley, T. J. *Paleoceanography* **2**, 449–478 (1987).
24. Hammer, C. U., Clausen, H. B. & Tauber, H. *Radiocarbon* **28**, 284–291 (1986).
25. Takahashi, T., Broecker, W. S. & Bainbridge, A. E. *Scope* **16** 159–200 (Wiley, New York, 1981).
26. Shackleton, N. J. & Chappell, J. *Nature* **324**, 137–140 (1986).
27. Lonsdale, P. *Deep Sea Res.* **24**, 1065–1101 (1977).
28. Emerson, S. *Am. geophys. Un. Monogr.* **32**, 78–88 (1984).
29. Pedersen, T. F. *Geology* **11**, 16–19 (1983).
30. Curry, W. & Lohmann, G. P. *Am. geophys. Un. Monogr.* **32**, 285–302 (1984).

Superheating, melting and vitrification through decompression of high-pressure minerals

Pascal Richet

Institut de Physique du Globe, 4, place Jussieu, 75232 Paris Cedex 05, France

A noteworthy feature of some high-pressure silicate minerals is their ready vitrification, either through moderate heating at 1 bar^{1–5}, as for jadeite ($\text{NaAlSi}_2\text{O}_6$) or stishovite (SiO_2), or through room-temperature decompression⁶, as for calcium silicate perovskite (CaSiO_3). Little attention seems to have been paid to the actual significance of such transitions, and here I argue that these simply represent crystal fusion. The unusual aspect is that the temperature of fusion at low pressures is below the temperature range of the glass–liquid transition. Through a different, thermochemical line of evidence, this extends an overlooked argument outlined by Jeanloz⁷ following the report of the glass/ice I transition by Mishima *et al.*⁸. The similarities and differences with this transition will be pointed out, as will some theoretical and geophysical implications.

To avoid ambiguity, the terms vitrification and melting will be used here for the experimentally observed fusions below and above the glass transition, respectively. For the purposes of the present discussion, the glass transition range (that is, the temperature interval in which the glass transition can take place under laboratory cooling rates) is restricted enough that it can be approximated as a single curve in the pressure–temperature (P – T) plane (Fig. 1a). Suppose that at some pressure P_i the melting curve of a mineral intersects the glass-transition curve of its amorphous phase (Fig. 1a). In principle, the mineral can bypass the liquid state and directly vitrify if brought from its

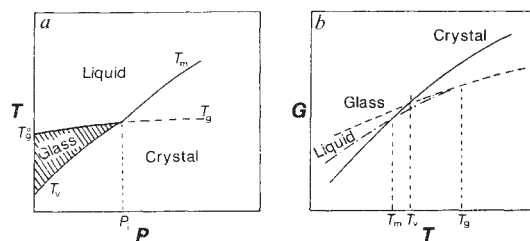


Fig. 1 a, Schematic 'phase diagram' when the fusion curve (T_m = melting, T_v = vitrification) intersects (at $P = P_i$) the glass transition range, approximated by a single curve, T_g , in the P – T plane. b, Gibbs free energies of formation (G) of the crystal, liquid and glass phases at $P = P_i$.

stability field to the shaded P – T area of Fig. 1a. (Note that at P_i the vitrification curve has a higher slope than the melting curve, because the Gibbs free energy of the glass must intersect that of the crystal at a higher temperature T_v than does the hypothetical supercooled liquid at T_m (see Fig. 1b), but such kinks should be minor, as suggested in Fig. 2 by the slight changes in the ΔG curves at T_g .)

For the SiO_2 polymorphs one readily calculates strong decreases in the 1-bar temperature of fusion (T_f^0) with increasing atomic packing, showing that the fusion scheme shown in Fig. 1a indeed applies for stishovite and also coesite, the other high-pressure form. The stable melting is that of cristobalite, at 2,000 K. For quartz, one finds⁹ that ΔG , the Gibbs free energy difference between crystal and liquid, is 0 at $T_f^0 = 1,700$ K (Fig. 2). This equilibrium melting point agrees with experimental observations of the metastable fusion of β -quartz¹⁰ and, although 300 K lower than that of cristobalite, such a temperature is still well above the glass transition of SiO_2 at $\sim 1,480$ K (ref. 9). For coesite and stishovite, similar calculations can be done with available heat capacities below¹¹ and above¹²

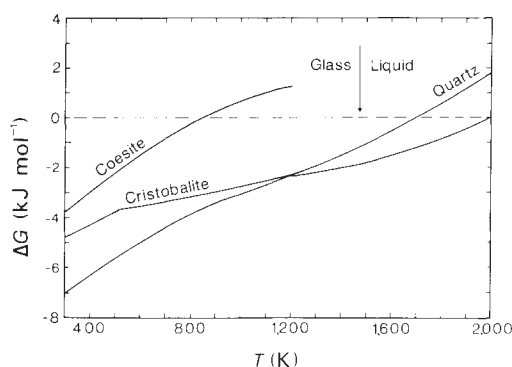


Fig. 2 Gibbs free energy differences (ΔG) between crystals and liquid (above 1,480 K) or glassy (below 1,480 K) SiO_2 (calculated, see ref. 9). Stishovite is not included because its ΔG is off the scale, with values ranging from 60 to 44 kJ mol^{-1} between 1,000 and 0 K. Note that the changes in the slopes of the curves at T_g are slight, because of partial cancellation of the effects of the C_p variations at T_g on the enthalpy and entropy terms. In particular, these changes are smaller than those resulting from the enthalpy and C_p changes at the α - β transitions in quartz (at 850 K) and cristobalite (at 525 K).

room temperature and enthalpies of formation¹³. The difference between these minerals and quartz is that both coesite and stishovite are unstable with respect to liquid SiO_2 at any temperature above, or even within, the glass transition interval of SiO_2 (Fig. 2). One may extend these calculations to the glass range⁹ and one finds for coesite $\Delta G = 0$ at $T_f^0 = 875$ K (Fig. 2). (The error margin of <200 K is mainly determined by the uncertainties on the enthalpy of formation of the mineral, and not by the influence of the thermal history¹⁴ on the Gibbs free energy of the glass.) Such a fusion temperature is definitely below the glass transition and, if there were no kinetic limitations, at 1 bar coesite could thus bypass the liquid state on heating and transform directly to a glass at temperatures between 875 K and $T_g = 1,480$ K. For stishovite, this could happen from 0 K because this mineral is unstable with respect to SiO_2 glass at any temperature, having still the enormous Gibbs free energy of vitrification of -44 kJ mol^{-1} at 0 K.

The fusion of jadeite cannot be studied in the same way because, in contrast to that of liquid SiO_2 , the absolute entropy of liquid jadeite cannot be obtained from the entropy of melting of a low-pressure polymorph. Jadeite melts congruently above 34 kbar and, from its melting curve determined up to 60 kbar^{15,16}, T_f^0 can be calculated through calculations of the 1-bar slope of the melting curve, $(dT/dP)^0$, as follows. In the Clausius-Clapeyron equation, $dT/dP = \Delta V/\Delta S$, the volume of fusion ΔV is readily calculated as a function of any assumed T_f^0 from thermal expansion^{17,18} and molar volume^{18,19} data for the crystal and the liquid. Likewise, as done in ref. 20, the entropy of melting ΔS can be derived from heat capacity^{18,20} and enthalpy of solution²¹ data as a function of T_f^0 . In this way, one obtains an initial Clapeyron slope that increases only slightly with the assumed T_f^0 (dashed lines in Fig. 3). On the other hand, if one arbitrarily fixes T_f^0 , then $(dT/dP)^0$ can be independently obtained from empirical melting-curve equations fitted to the high-pressure melting data. The Simon equation is widely used for silicates (see, for example, ref. 22), and here we use it in a purely empirical way for interpolation purposes. The results of calculations done with this equation (solid curves in Fig. 3) show that these $(dT/dP)^0$ strongly depend on the fixed T_f^0 , giving sharp intersections with the Clausius-Clapeyron values at $T_f^0 = 970 \pm 20$ K. This equilibrium 1-bar melting temperature is 100–150 K below the glass-transition range²⁰, consistent with experimental observations of the decomposition of jadeite to an amorphous phase at 1,000–1,100 K^{1-3} .

The possibility of obtaining stishovite and other high-pressure minerals in the P - T stability field of the glass does not contradict

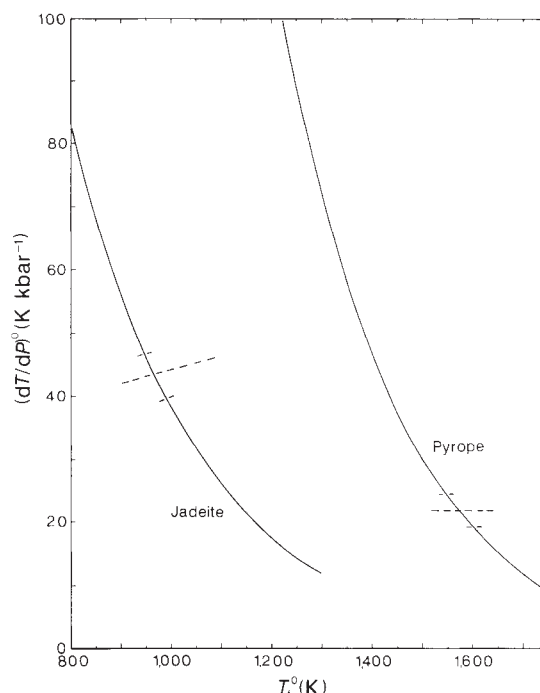


Fig. 3 Initial slopes of the melting curves of jadeite and pyrope. Dashed lines, values calculated with the Clausius-Clapeyron equation; solid lines, values derived from the Simon equation, $P = a [(T/T_f^0)^b - 1]$, fitted to the experimental melting data^{15,16,31,32}. The coefficients of the Simon equation obtained for jadeite are $a = 7.45$ kbar and $b = 3.036$, with $T_f^0 = 970$ K; for pyrope, $a = 14.96$ kbar and $b = 4.757$, with $T_f^0 = 1,570$ K. The error bars reported in the text for T_f^0 were obtained from the short dashed lines, which delineate the range of $dT/dP = \Delta V/\Delta S$ obtained when the uncertainties on the ΔV and ΔS data are taken into account.

the above thermodynamic analysis, which says nothing about the vitrification kinetics. In fact, the lowest temperature from which vitrification may be experimentally observed is an upper bound for T_f^0 . The difference between this temperature and T_f^0 is the superheating interval of the crystal, where vitrification occurs irreversibly. At 1 bar, superheating is restricted to ~ 100 K for jadeite, which is similar to the superheating reported for quartz²³ and high albite²⁴ at higher temperatures. The superheating interval of coesite seems much wider, of the order of 400–500 K, vitrification having been reported at 1,370 K but not at 1,270 K (ref. 5). Stishovite should vitrify at 0 K but begins to do so very slowly just above room temperature, and is capable of undergoing considerable superheating when kept at temperatures of more than 1,000 K for a few minutes^{4,5}.

Vitrification usually derives from the cooling of liquids. In the case of water, however, Mishima *et al.*⁸ obtained a glass through compression of ice I, which has a negative dT/dP , above its (extrapolated) melting pressure in a P - T range where ice VI is stable. The common feature with the present evidence for coesite, stishovite and jadeite is that, as previously noted⁸, glass formation below the glass transition results from the inability of metastable crystals to transform to the stable polymorph, or mineral assemblage, in a certain, low temperature range. Such crystals should either superheat or vitrify and, from their results on water, Mishima *et al.*⁸ suggested that glasses might be generally formed when metastable compounds with negative volumes of fusion are compressed.

A priori, there is no reason why vitrification should be restricted to compounds with negative ΔV s of fusion, with the sole difference that crystals with positive ΔV s must be decompressed to vitrify. Practically, this favours such a behaviour for compounds stable under extreme pressures. Thus the present

evidence is a symmetrical complement to the argument of Mishima *et al.*, with the difference that superheating seems much more restricted for crystals like ice than for silicates. One might go a step further by considering the thermodynamically possible situation that stable, and not only metastable, compounds also have a melting point below the glass transition range of the amorphous phase. Such compounds are not known, but another silicate, high albite ($\text{NaAlSi}_3\text{O}_8$), could constitute the closest approximation. It has a melting point somewhere below 1,400 K (ref. 25), at which the viscosity of the liquid is only a few orders of magnitude less than the 10^{13} poise at the glass transition $T_g > 1,100$ K (ref. 20), whence the reported extreme difficulty of measuring T_f^0 (refs 3, 25).

From a thermochemical standpoint, glass formation below the glass transition raises some questions. Are the enthalpies of glasses formed at different temperatures a function of the vitrification temperature? If different, is the concept of fictive temperature, used to define the thermodynamic state (see, for example, ref. 14) still operationally valid and how is it related to that of the fictive pressure, which can be used for characterizing vitrification through decompression at constant temperature? Glass samples big enough for calorimetric or spectroscopic measurements could be prepared from jadeite or natural stishovite to investigate these and other problems, such as the temperature dependence of the theoretically important²⁶, but unobservable heat capacity of supercooled liquids below the glass transition; the existence of possible glass-glass transitions as reported in the case of water²⁷; or the relationship between these glasses and 'usual' glasses permanently densified at high pressures.

Knowledge of the fusion curves of silicate perovskites, thought to be major constituents of the Earth's mantle, and of other high-pressure minerals, is of particular interest in geophysics. Determining the temperature of the onset of vitrification at 1 bar as well as at pressure would give upper bounds for the fusion curves and put constraints on volume changes of fusion as a function of pressure, as well as on 1-bar thermochemical properties of minerals, those of the glasses being known independently¹⁴. This would be useful for calcium silicate perovskite, for example, which vitrifies when quenched to 1 bar (ref. 6), suggesting that its fusion curve is very steep up to ~200 kbar where its high-temperature stability indicates high melting points. For (Mg, Fe)SiO₃ perovskite, an almost constant melting temperature of 3,000–3,200 K between 200 and 600 kbar has been reported²⁸. By analogy with calcium silicate perovskite, its fusion temperature should decrease dramatically towards 1 bar, which could be checked through observation of primary vitrification before the eventual formation of pyroxene observed at ~600 °C (ref. 29).

Dense minerals with sixfold-coordinated silicon are prime candidates for ready low-pressure vitrification. To conform to Fig. 1a, the fusion curve must have a steep slope at low pressure, which can result from a high 1-bar ΔV of fusion. Stishovite and calcium silicate perovskite illustrate this point, but the present analysis shows that this structural feature could be a sufficient, although not a necessary condition. Both coesite, made up of a compact arrangement of SiO₄ tetrahedra, and jadeite, whose structure rests on SiO₄ tetrahedra and AlO₆ octahedra, are prone to vitrification. But the presence of octahedral aluminium in high-pressure silicates is not a sufficient criterion. Pyrope garnet (Mg₃Al₂Si₃O₁₂) has a pressure range of stability similar to that of jadeite and is also made up of tetrahedral silicon and octahedral aluminium. The fusion of this garnet, however, does not obey Fig. 1a, because one obtains a 1-bar melting point of $1,570 \pm 30$ K (Fig. 3), well above the calorimetric T_g of 1,070 K (ref. 30), from (dT/dP) calculations done as described above for jadeite with the relevant volume^{17–19}, thermochemical³⁰ and high-pressure melting^{31–32} data.

Helpful comments on the manuscript were made by Y. Bottinga, F. Guyot, J. Peyronneau, J. P. Poirier and B. J. Wood.

Received 20 March; accepted 17 November, 1987.

- Greig, J. W. & Barth, T. F. W. *Am. J. Sci.* **35A**, 93–112 (1938).
- Yoder, H. S. Jr *Am. J. Sci.* **248**, 225–248; 312–334 (1950).
- Schairer, J. F. & Bowen, N. L. *Am. J. Sci.* **254**, 129–195 (1956).
- Skinner, B. J. & Fahey, J. J. *J. geophys. Res.* **68**, 5595–5604 (1963).
- Dachille, F., Zeto, R. J. & Roy, R. *Science* **140**, 991–993 (1963).
- Liu, L. G. & Ringwood, A. E. *Earth planet. Sci. Lett.* **28**, 209–211 (1975).
- Jeanloz, R. *EOS* **65**, 1245–1246 (1984).
- Mishima, O., Calvert, L. D. & Whalley, E. *Nature* **310**, 393–395 (1984).
- Richet, P., Bottinga, Y., Denielou, L., Petitot, J.-P. & Téqui, C. *Geochim. cosmochim. Acta* **46**, 2639–2658 (1982).
- Mackenzie, J. D. *J. Am. Ceram. Soc.* **43**, 615–620 (1960).
- Holm, J. L., Kleppa, O. J. & Westrum, E. F. Jr *Geochim. cosmochim. Acta* **31**, 2289–2307 (1967).
- Watanabe, H. in *High-pressure Research in Geophysics* (eds Akimoto, S. & Manghnani, M. H.) 441–464 (Center for Academic Publications, Tokyo, 1982).
- Akaogi, M. & Navrotsky, A. *Phys. Earth planet. Inter.* **36**, 124–134 (1984).
- Richet, P. & Bottinga, Y. *Rev. Geophys.* **24**, 1–25 (1986).
- Bell, P. M. & Roseboom, E. H. Jr *Mineralog. Soc. Am. spec. Pap.* **2**, 151–161 (1969).
- Williams, D. W. & Kennedy, G. C. *Am. J. Sci.* **269**, 481–488 (1970).
- Skinner, B. J. *Mem. geol. Soc. Am.* **97**, 75–96 (1966).
- Bottinga, Y., Weill, D. F. & Richet, P. *Geochim. cosmochim. Acta* **46**, 909–919 (1982).
- Robie, R. A., Hemingway, B. S. & Fisher, J. R. *Bull. U.S. geol. Surv.* No. **1452** (1978).
- Richet, P. & Bottinga, Y. *Geochim. cosmochim. Acta* **44**, 453–470 (1984).
- Navrotsky, A., Hon, R., Weill, D. F. & Henry, D. F. *Geochim. cosmochim. Acta* **44**, 1409–1423 (1980).
- Bottinga, Y. *Earth planet. Sci. Lett.* **74**, 350–360 (1985).
- Ainslie, N. G., Mackenzie, J. D. & Turnbull, D. *J. phys. Chem.* **65**, 1718–1724 (1961).
- Dietz, E. D., Baak, T. & Blau, H. H. *Z. Kristallogr. Kristallgeom.* **132**, 340–360 (1970).
- Boettcher, A. L., Burnham, C. W., Windom, K. E. & Bohlen, S. R. *J. Geol.* **90**, 127–138 (1982).
- Kauzmann, W. *Chem. Rev.* **43**, 219–256 (1948).
- Mishima, O., Calvert, L. D. & Whalley, E. *Nature* **314**, 76–78 (1985).
- Heinz, D. L. & Jeanloz, R. *Abstr. U.S.-Japan Seminar on High-Pressure Research Applications in Geophysics and Geochemistry*, 16 (University of Hawaii, Honolulu, 1986).
- Knittle, E. & Jeanloz, R. *Nature* **319**, 214–216 (1986).
- Richet, P. & Bottinga, Y. *Earth planet. Sci. Lett.* **67**, 415–432 (1984).
- Boyd, F. R. & England, J. L. *Yb. Carnegie Instn. Wash.* **61**, 107–112 (1962).
- Ohtani, E., Irifune, T. & Fujino, K. *Nature* **294**, 62–64 (1981).

Laboratory models of Hawaiian and Strombolian eruptions

Claude Jaupart & Sylvie Vergnolle

Laboratoire de Dynamique des Processus Géologiques,
Université Paris 7 et Institut de Physique du Globe de Paris,
4 place Jussieu, 75252 Paris Cedex 05, France

Basaltic eruptions are often characterized by cyclic changes of activity. At Hawaii, periods of continuous fountaining alternate with much longer periods of effusive outflow^{1,2}. In Strombolian eruptions, activity proceeds through intermittent discrete bursts^{2–5}. We report laboratory experiments that simulate the degassing process in basaltic eruptions. Gas bubbles are generated at the bottom of a tank filled with viscous liquid and topped by a small open conduit. The bubbles rise and accumulate at the roof in a foam layer whose thickness increases. At a critical thickness the bubbles coalesce and the foam collapses, generating gas pockets whose size depends on liquid viscosity and surface tension. At low viscosity a single large gas pocket is formed, which flows into the conduit. This erupts in an annular flow configuration where a central jet expels the liquid films that wet the conduit walls⁶. At higher viscosity many smaller pockets are formed, which rise as slugs and burst out intermittently at the vent. The experiments imply that the presence of constrictions in the chamber and conduits plays a major role in determining eruption behaviour.

There is evidence that, in basaltic volcanoes, magma can degas at depth before eruption⁷. For example, during eruptions at Etna, magmatic gases are continuously exsolved in a reservoir and have a large residence time there⁸. This indicates that magma is degassing in the chamber feeding the eruption. To investigate the consequences of this we have studied the behaviour of gas bubbles rising close to a constriction. Such constrictions could be at the top of a large magma chamber as on Hawaii or along the axis of a feeder dyke where activity has been focused to a central vent as in most basaltic fissure eruptions. Figure 1 shows the experimental tank: at the top there is a long conduit; at the bottom a set of capillary tubes release gas bubbles over the

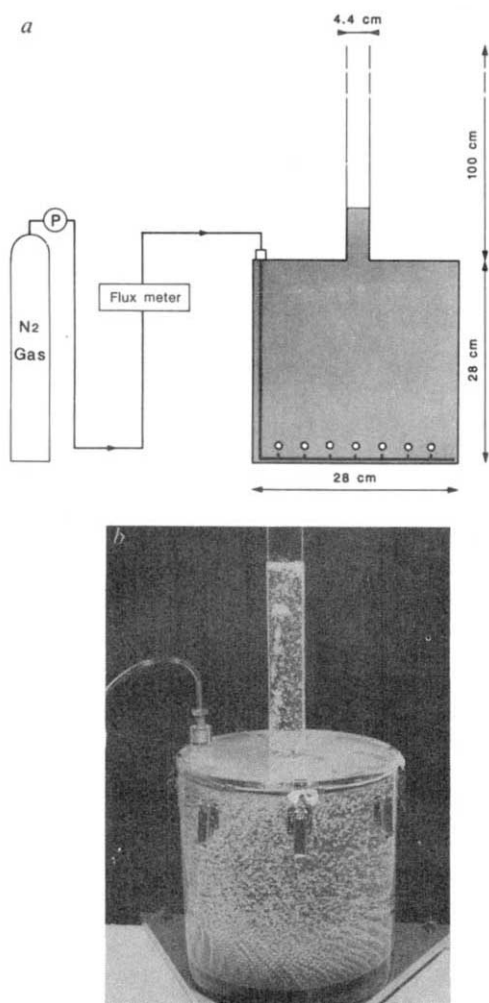


Fig. 1 *a*, Experimental set-up to simulate magma degassing at a constriction (not to scale). Both the gas flux and bubble size are monitored. *b*, Typical experiment (with 0.1 Pa s liquid viscosity). Flow patterns in the tank interior are complex but are not essential to the phenomena under study. The gas bubbles accumulate at the roof in a foam layer.

whole surface. Both the overall gas flux and bubble size are monitored. We experimented with different numbers of capillary tubes (185 and 38), corresponding to different bubble separation. We also used conduits of different diameters (4.4 and 9.4 cm), representing different roof/conduit area ratios. The capillary tubes and gas fluxes were selected to ensure a small range of bubble radii throughout the experiments: from 0.9 to 2 mm. The gas flux was varied from 200 to 2,000 cm³ min⁻¹. The working fluids were silicone oils with viscosities from 0.1 to 5 Pa s.

The basic physical situation is that the bubbles rise and accumulate at the tank roof in a foam layer (Fig. 1*b*). The phenomena observed depend on the liquid viscosity, and are not sensitive to either the bubble separation or the roof/conduit area ratio. The role of the gas flux will be discussed last.

At low viscosity (0.1 Pa s), we observed cyclic changes of regime. First the foam accumulated and liquid was expelled in the conduit to compensate for the increase of gas volume in the tank (Fig. 2*a*). When the foam thickness reached a critical value, the bubbles coalesced. The whole foam collapsed almost instantaneously into a single large gas pocket. This erupted into the conduit (Fig. 2*b*), leading to annular flow with liquid films along the walls (Fig. 2*c*). Once the pocket had erupted, the liquid column in the conduit fell (Fig. 2*d*) because the volume of gas trapped at the roof was gone. The cycle was repeated with striking regularity. During foam accumulation, part of the foam

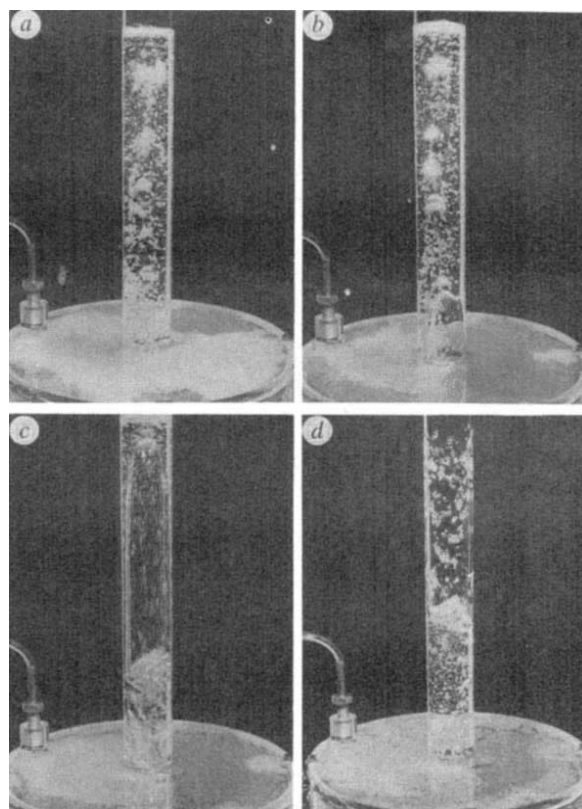


Fig. 2 Time sequence of events at low viscosity (0.1 Pa s), at 0.5 s intervals. *a*, The foam accumulates at the roof and the level of the liquid column in the conduit rises. *b*, The foam collapses into a single large gas pocket, which starts erupting: the nose of the pocket can be seen at the conduit entrance. *c*, The gas pocket erupts in an annular flow configuration with a central gas jet and liquid films along the conduit walls. *d*, The gas pocket has erupted. Note the large drop of liquid level in the conduit.

flowed into the conduit, sometimes coalescing into small gas pockets at the conduit entrance. These rose and burst out at the top of the liquid column.

We compare these observations to a report from the 1969–71 Mauna Ulu eruption of Kilauea volcano¹. During the first six months, fire fountains lasting from 5 h to 3 days were interspersed with weeks of quiet effusive activity. This cycle is reproduced in the experiments: a long period of gas accumulation resulting in slow outflow followed by a much shorter phase of vigorous activity. In the field, the beginning of fire fountaining is such that lava wells out of the vent, the extrusion rapidly grows in height and develops into a fountain. This is consistent with a large gas pocket displacing liquid upwards as it rises and decompresses in the liquid-filled conduit (Fig. 2*c*). The onset of fountaining corresponds to the top of the pocket reaching the vent. Fire fountaining ends abruptly. At that time the top of the lava column lies at a depth of more than about 100 m: the conduit has been emptied and some erupted lava pours back into it. We too observed such a fall in liquid level (Fig. 2*d*). Between two fire fountains, lava in the conduit follows a cycle of rise and fall termed 'gas-piston activity' associated with the rise and bursting of a gas pocket. This is also reproduced.

At higher viscosity (1–5 Pa s) there was a different behaviour. Rather than going through a cycle of accumulation and collapse, the foam maintained an approximately constant thickness (close to the critical thickness determined above), with a complex distribution of coalesced gas pockets (Fig. 3*a*). These pockets flowed into the conduit, rising as gas slugs that burst out intermittently at the top (Fig. 3*b,c*). The slugs were quite uniform in

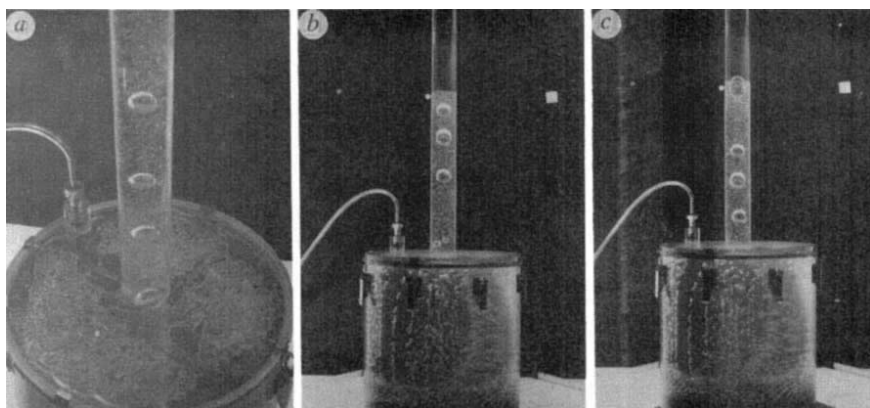


Fig. 3 Behaviour at higher viscosity (1 Pa s). *a*, Close-up of the tank roof, showing gas pockets of various sizes resulting from bubble coalescence. The largest gas pockets flow into the conduit. *b*, Intermittent flow in the conduit. Note that the size of the gas slugs is constant and compare it with the bubble size in the tank. *c*, A gas slug bursting at the top of the liquid column. Note that variations of liquid level are negligible.

size, almost as wide as the conduit, and expelled liquid clots upon bursting. These features are strongly reminiscent of Strombolian activity²⁻⁵.

We now turn to the physics of the experiments. Consider first the foam. As it accumulates at the roof, it flows horizontally towards the conduit. At low gas flux, a steady state can be reached such that this flow balances the input of bubbles from below, and hence the foam layer adopts an equilibrium shape. The foam behaves as a liquid of effective viscosity $\mu = \mu_1 \cdot f(\epsilon)$, where μ_1 is the liquid viscosity and f a function of the gas volume fraction ϵ (ref. 9). The governing equations are similar to those for viscous currents^{10,11}. We find that, for given roof and conduit, the maximum foam thickness scales with $h_1 = \{Q\nu/g\}^{1/4}$, where Q is the gas flux, ν the foam kinematic viscosity and g the acceleration of gravity. We focus next on bubbles at the top of the foam. Buoyancy acts to flatten them against the tank roof (see ref. 12). As the foam thickness increases, so does the total buoyancy force applied on these bubbles. Eventually, surface tension can no longer sustain this force and the bubbles must burst out and coalesce. A balance between buoyancy and surface tension yields that the critical thickness scales with $h_2 = \sigma/(\epsilon\rho bg)$, where σ is the surface tension coefficient, ρ the liquid density and b the bubble radius. If the foam thickness is smaller than this critical thickness, no coalescence is possible and bubbly flow is observed in the conduit. In short, for bubble coalescence to occur, the gas flux must exceed a critical value.

Consider next the generation of gas pockets. Once initiated, the phenomenon is unstable. As two adjacent bubbles coalesce, the gas inside must expand to restore equilibrium with surface tension¹³. This acts to thin the liquid film surrounding the incipient pocket, and hence to rupture it, thus adding new bubbles to the pocket. This process is driven by capillary pressure and occurs over a characteristic time $(\mu_1 \cdot b)/\sigma$ (the 'growth time'). As the growing gas pocket impinges on the conduit entrance, it escapes. Thus, the pocket is transported along with the overall flow and grows while resident in the foam layer. If the growth time is less than the residence time, the whole foam collapses into a single gas pocket. At higher viscosity, the growth time is close to the residence time, and gas slugs are generated. The ratio of the two characteristic times defines a dimensionless number T . The gas slugs rise in an inertia-dominated regime and their velocity scales with $\sqrt{(gD)}$, where D is the conduit diameter (ref. 14, p. 312).

These considerations are consistent with the experimental results. The critical fact is that the gas pocket volume does not depend on the gas flux, and is a function of T . The time between two successive gas pockets is simply the time needed to re-establish the critical foam thickness. It is thus inversely proportional to the gas flux. It is also proportional to the pocket volume, and hence is much smaller at high viscosity: for a 1,000-cm³ min⁻¹ gas flux, the 'fire fountains' and 'explosions' of Figs 2 and 3 are repeated every 20 s and 2 s respectively.

Departures from the sequence of events described must be

expected owing to the complex shape of a magma chamber roof. In its barest form, our proposal is that fire fountains and explosions are both due to bubble coalescence at a constriction, and can be described as periodic 'bursts' of different intensity. The longer the periodicity, the longer the duration of the burst. At Stromboli the periodicity and duration are tens of minutes and seconds respectively³, whereas they are weeks and hours at Hawaii¹. One could interpret these differences in terms of variations in catchment area for accumulating gas, degassing rate or lava viscosity. We have emphasized the gas phase, but liquid is also erupted in the experiments. Annular flow drives upwards the liquid films that wet the conduit walls (a phenomenon known as 'flooding'; ref. 14, p. 336). Gas slugs expel fragments of their envelope. From photographs we find that the proportion of liquid to gas is smaller in the latter case. This proportion is about 10^{-2} at Hawaii⁶ and 10^{-5} at Stromboli³.

For the model the key parameter and unknown is the bubble size in the chamber. In a detailed analysis to be submitted elsewhere, we calculate the bubble radius required to fit the Mauna Ulu gas flux. We conclude by emphasizing a few facts about the model. It reproduces surface observations of certain Hawaiian and Strombolian eruptions. It is consistent with the inference of Chouet *et al.*³ at Stromboli that the extreme volume ratio of gas to melt measured there requires volatile concentration at depth. Lastly, at Hawaii, the volume fluxes of lava during effusive activity and fire fountains differ by at least a factor of 10 (refs 15, 16). Both regimes are close to steady state because their characteristics do not vary over a period exceeding the rise-time from chamber to vent¹⁷. Thus they reflect different conditions at the conduit entrance. Cyclic changes are needed in the magma chamber itself, which are accounted for by our mechanism.

We thank Stephen Tait and Lionel Wilson for comments. Herbert Huppert and Stephen Sparks contributed stimulating criticisms.

Received 3 June; accepted 2 October 1987.

- Swanson, D. A., Duffield, W. A., Jackson, D. B. & Peterson, D. W. *U.S. geol. Surv. prof. Pap.* 1056 (1979).
- Williams, H. & McBirney, A. R. *Volcanology* (Freeman, San Francisco, 1979).
- Chouet, B., Hamisevicz, N. & McGetchin, T. R. *J. geophys. Res.* **79**, 4961-4975 (1974).
- Blackburn, E. A., Wilson, L. & Sparks, R. S. J. *J. geol. Soc. Lond.* **132**, 429-440 (1976).
- Wilson, L. *J. Volcan. geotherm. Res.* **8**, 297-313 (1980).
- Vergnolle, S. & Jaupart, C. *J. geophys. Res.* **91**, 12842-12860 (1986).
- Greenland, L. P. in *Prof. Pap. U.S. geol. Surv.* **1350**, 781-790 (1987).
- Lambert, G., le Cloarec, M. F., Ardouin, B. & le Rouley, J. C. *Earth planet. Sci. Lett.* **76**, 185-192 (1986).
- Sibree, J. O. *Trans. Faraday Soc.* **30**, 325-331 (1934).
- Huppert, H. E. *J. Fluid Mech.* **121**, 43-58 (1982).
- Huppert, H. E., Shepherd, J. B., Sigurdsson, H. & Sparks, R. S. J. *J. Volcan. geotherm. Res.* **14**, 199-222 (1982).
- Lee, J. C. & Hodgson, T. D. *Chem. Engng Sci.* **23**, 1375-1397 (1968).
- Bikerman, J. J. *Foams* (Springer, Berlin, 1973).
- Wallis, G. B. *One-dimensional two-phase flow* (McGraw-Hill, New York, 1969).
- Swanson, D. A., Jackson, D. B., Duffield, W. A. & Peterson, D. W. *Geotimes* **16**, 12-16 (1971).
- Swanson, D. A. *Geol. Soc. Am. Bull.* **84**, 615-626 (1973).
- Wilson, L. & Head, J. W. III *J. geophys. Res.* **86**, 2971-3001 (1981).

Microscale isotopic zoning in calcite and graphite crystals in marble

Hideki Wada

Institute of Geosciences, Shizuoka University, 836 Ohya Shizuoka, 422 Japan

Chemical and isotopic zoning in minerals is a clear indication of disequilibrium and provides an important key for elucidating metamorphic processes. Here I present carbon and oxygen isotope data from calcite and graphite crystals in a marble from the Hida metamorphic belt, central Japan, showing isotopic zoning within individual crystals. Within 100 μm of the calcite-graphite grain boundary the ^{13}C contents gradually decrease from the constant values of the crystal core to values $\sim 1\%$ lower; ^{18}O contents in calcite decrease by up to 7% within 200–300 μm of the boundary. At a calcite-calcite grain boundary, similar patterns of ^{13}C and ^{18}O depletions were found, whereas the ^{13}C contents of graphites remained constant except in a narrow zone at the calcite contact. These isotopic patterns are interpreted as reflecting isotopic exchange with metamorphic fluids that were relatively depleted in ^{13}C and ^{18}O . These results strongly suggest that the most accurate geothermometry will be obtained if only the cores of crystals are used, and that isotopic zoning may provide another way to elucidate polymetamorphic processes.

It is well known that many minerals are zoned in chemical composition. In the case of stable isotopes, minerals are known to be isotopically homogeneous on scales exceeding several hundred metres¹. On the other hand, heterogeneity on a centimetre scale is also known in some cases². Thus, stable isotope zonation in mineral crystals may occur. Garlick and Epstein reported³ that isotopic zonation in two metamorphic garnets was negligibly small. Knowledge of the grain-to-grain isotopic composition is one criterion used to test for attainment of equilibrium.

The empirically established carbon isotope geothermometer for the calcite-graphite system has proved useful when applied to intermediate-to-high grades of metamorphisms^{4–6}. However, it has been pointed out that, in intermediate-to-low grades of metamorphism, calcite and graphite do not always attain equilibrium^{4,5,7,8} and therefore the application of the carbon isotope geothermometer is restricted. In the case of some high-grade metamorphic rocks, polycrystalline graphite was found to be heterogeneous in carbon isotopic composition^{4,9,10}. Microscale determinations of the isotopic composition of calcite and graphite are required to clarify the isotopic behaviour during metamorphism. Recent developments of the laser method¹¹ or ion-probe micro analyser¹² could provide powerful tools for determining the isotopic composition of ultra small samples, but their resolution is still not as high as can be achieved with conventional mass spectrometry.

The result of carbon and oxygen isotopic analyses on three traverses of a calcite-graphite-calcite grain boundary (Fig. 1) are shown in Fig. 2. In all three traverses calcite showed $\delta^{13}\text{C}$ values (where $\delta^{13}\text{C} = ((^{13}\text{C}/^{12}\text{C})_{\text{sample}} - (^{13}\text{C}/^{12}\text{C})_{\text{standard}}) / (^{13}\text{C}/^{12}\text{C})_{\text{standard}}$) of about +3.0‰ which gradually decreased to about +2.0‰ near the graphite contacts. For the graphite, $\delta^{13}\text{C}$ values seemed to be constant, except for a very narrow zone at both sides of the graphite crystal where $\delta^{13}\text{C}$ values decreased sharply. In the traverse B', $\delta^{13}\text{C}$ values of graphite coincided with that of the core part of the graphite in the A-B traverse.

Oxygen isotope behaviour contrasts to that of carbon isotopes: $\delta^{18}\text{O}$ values were constant at -4.7% in the core of the calcite crystal, while toward the grain boundary they gradually decreased by $\sim 4\%$. Figure 2 shows that two $\delta^{18}\text{O}$ values (-5.4 and -5.3%) for calcite sampled at 500- μm intervals toward the

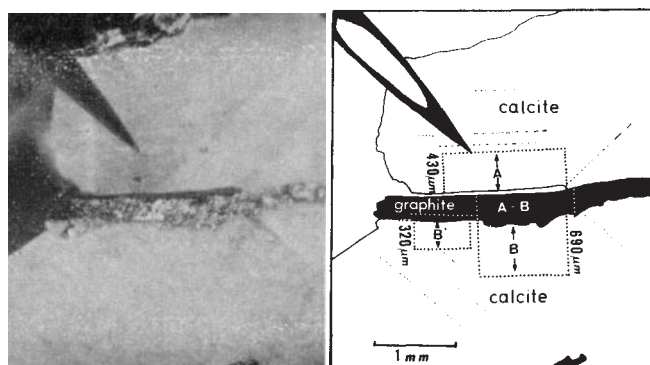


Fig. 1 Micro-sampling loci of the calcite and graphite crystals in marble collected from Wada-gawa area of the Hida metamorphic belt, central Japan. Three traverses (A, B and B') of calcite-graphite pairs were examined in one suite of calcite-graphite-calcite grain boundaries. These traverses cut perpendicular to the plane of calcite-graphite contact. The graphite crystal is about 3-mm long and 0.2-mm thick. An artificial crack can be seen in the grain boundary of traverse A. The marble studied consisted mainly of calcite with subordinate amounts of scaly graphite and silicates. The saccharoidal calcite crystals had grown to ~ 1 cm in size, while the graphite crystals, found mostly along the boundaries between calcite crystals, often consist of hexagonal flakes a few mm in diameter. Scaly graphitic marble is common in the Hida metamorphic belt, which is thought to have undergone high-grade polymetamorphism up to granulite facies^{14,17}.

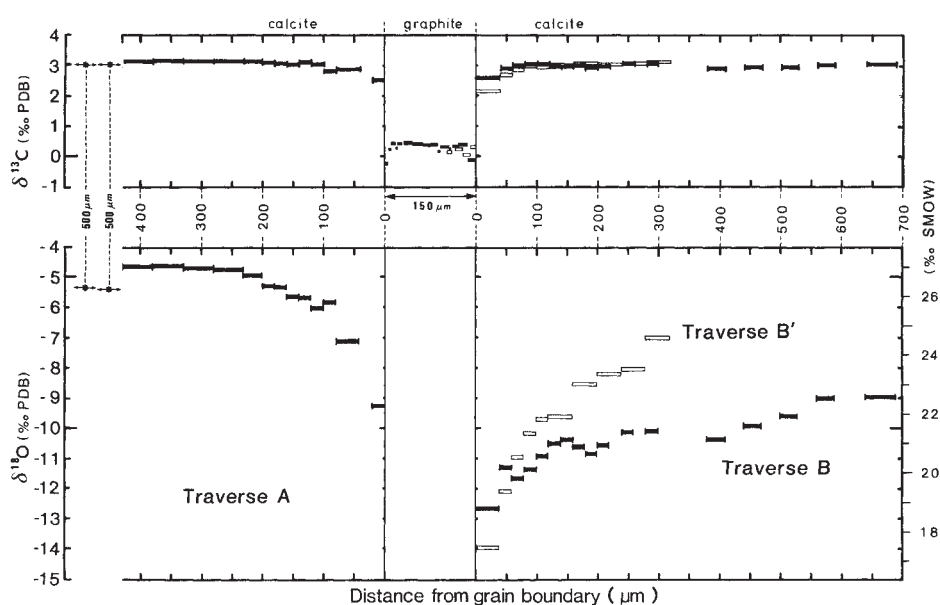
crystal core were slightly lower than the levelled value of -4.7% of traverse A. In traverse B, $\delta^{18}\text{O}$ values increased from -12.7 to -9.0% along the traverse toward the core from the boundary with fluctuations of amplitude 1%.

The cores of several thin sections of adjoining calcite crystals were analysed to test the isotopic homogeneity among cores of individual calcite crystals; the sampling loci and the isotopic results are shown in Fig. 3. The $\delta^{13}\text{C}$ values of cores of individual grains were constant to within $\pm 0.1\%$. The $\delta^{18}\text{O}$ values, however, varied from -5.5 to -6.2% , values which are slightly different from the value -5.3% of the traverse A core.

Another calcite crystal from a different part of the same sample was examined on a calcite-calcite grain boundary traverse. Figure 4 shows that the pattern of $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values of the calcite are similar to those of the calcite-graphite boundaries described above. A very low $\delta^{18}\text{O}$ value of -16.0% was measured at the immediate contact and increased toward the crystal core. Oxygen isotope fluctuation of up to 1.2‰ is likely in the core of the calcite crystal; however, the magnitude was much smaller than that found at the grain boundary.

As there are only trace amounts of silicates in the marble, decarbonation and dehydration processes cannot account alone for the ^{18}O depletion. An exchange reaction with low ^{18}O metamorphic fluids infiltrated during the last stage of metamorphism seems to be the only possible mechanism that could account for the marked ^{18}O depletion trend. In the traverses of calcite-graphite-calcite shown in Figs 1 and 3, the directions of cleavage in the calcite crystals differed on each side of the graphite. In traverse A, rhombohedral cleavage parallel to the knife-sharp contact surface with graphite was observed. In traverses B and B', the cleavage line of the calcite crystal obliquely intersected the graphite crystal boundary. The calcite and graphite contact of traverse B' is knife-sharp while the contact of traverse B is irregular. Small graphite crystals grew penetratively or intercalatively in the calcite crystal of traverse B. It is plausible that such grain boundaries act as channels for the metamorphic fluid. The irregularity of calcite-graphite boundaries and the existence of minor graphite inclusions are likely to relate to the irregular

Fig. 2 Isotopic results of calcite and graphite crystals of traverse A, B and B'. Data from the core part of the traverse A crystal are also plotted as asterisks. Sampling loci are illustrated in Fig. 1: $\delta^{13}\text{C}$ values of graphite A-B are plotted as solid rectangles; open rectangles show $\delta^{13}\text{C}$ values in graphite contact with B' calcite. To determine accurately carbon and oxygen isotopic compositions of 0.1- μmol samples of carbon dioxide, some modifications of the method were required^{10,18,19}. The key modifications are: (1) the micro-inlet system of the mass spectrometer was modified for small samples; (2) the pressure effect due to reducing sample pressure relative to the working standard was compensated for by giving appropriate offsets on the zero adjustments of the amplifiers for the masses 44, 45 and 46; (3) water vapour was frozen out using a cold trap filled with *n*-pentane/liquid nitrogen mixtures. The Finnigan MAT-250 mass-spectrometer was modified to allow analyses of samples as small as 2- μl CO_2 at STP. Graphite samples were combusted with vanadium pentoxide in Vycor glass tubes at 1,100 °C. Through this combustion procedure, ~ 0.1 μl of ^{13}C -depleted CO_2 evolved from the Vycor glass and mixed with the sample gas. Therefore, 10 μl or more of sample CO_2 was required to maintain a precision better than 0.1% in the determination of the $\delta^{13}\text{C}$ value of graphite. The calcite sample, contained in a small stainless steel cup, was dropped into a reaction vessel connected to the mass-septometer and containing concentrated phosphoric acid at 60 °C. Three traverses (traverse A, B and B') of calcite-graphite pairs were examined in one suite of calcite-graphite-calcite grain boundaries (shown in Fig. 1). The calcite crystal was trimmed into a prism by a handheld diamond saw. The cross-sections of the calcite prisms along the three traverses A, B and B' were 1.7, 1.5 and 0.65 mm^2 , respectively. The prisms were mounted in a block of ice on the sample holder of an electro-freezing microtome and the calcite crystal was sliced together with the ice block at intervals of 20 μm or more. Sampling the graphite crystal involved cutting a piece about 1 mm^2 in cross-section and repeatedly exfoliating perpendicular to its *c*-axis with adhesive tape. Graphite flakes of appropriate thickness were selected from the tape and washed with toluene. The graphite flakes were then scooped up by molybdenum mesh and combusted in a Vycor glass tube. The average thickness of a graphite flake was calculated from the volume of CO_2 evolved and the area of the sample flake.



pattern of $\delta^{18}\text{O}$ values of traverse B. Unlike the calcite-graphite boundaries which are generally smooth, calcite-calcite boundaries are often irregular. Furthermore, at the calcite-calcite boundary, a band of mosaic grains of calcite crystals a few tens of μm in width is often observed under the microscope. The low $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values of calcite within 40 μm of the boundary, as shown in Fig. 4, indicate that these small calcite grains should be considerably depleted in ^{13}C and ^{18}O .

The $\delta^{18}\text{O}$ values (-8.1 to -9.5%) of the calcite crystal core of traverse C do not agree well with that of traverse A (-5.6 to -5.3%). This indicates an isotopic heterogeneity of up to 4% in hand-size samples. The discrepancy between the $\delta^{18}\text{O}$ values of the calcite crystals on either side of the graphite crystal (traverse A and B) may be caused by differences in permeability of the two boundaries. The homogeneity in the core of calcite crystals indicates that carbon isotopic equilibrium had been attained in them, while the fact that $\delta^{13}\text{C}$ values decrease in both calcite and graphite toward grain boundaries indicates that late-stage metamorphic fluids were relatively depleted in ^{13}C as well as in ^{18}O .

The marble studied consisted of coarse grained (1–10 mm) saccharoidal calcite crystals grown during prograde metamorphism in which small calcite grains grew to a larger size to reduce the surface free energy. Small amounts of metamorphic fluid may affect such a recrystallization process. Carbon dioxide in the metamorphic fluid should be slightly enriched in ^{13}C as a result of equilibration with calcite¹³ at temperatures higher than 200 °C. Therefore, the ^{13}C depletion in the margins of calcite cannot be explained by exchange with CO_2 in metamorphic fluid during the temperature drop from granulite to amphibolite facies. In the case of inter-mineral diffusion, opposite trends in $\delta^{13}\text{C}$ values of calcite and graphite crystals

to that observed should be expected. Kinetic fractionation during recrystallization may cause depletion, but a quantitative discussion is not possible at this stage.

Another possible explanation assumes a two stage metamorphism. The margin of the calcite crystal was infiltrated by metamorphic fluids, containing a ^{13}C -depleted component (methane or higher hydrocarbons). It is believed that this retrogressive metamorphism up to amphibolite facies followed the peak metamorphism of granulite facies¹⁴.

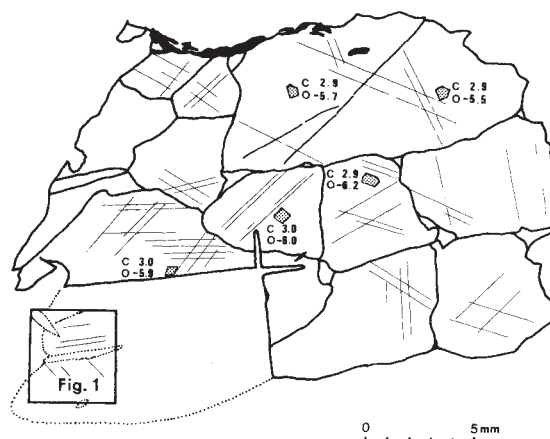


Fig. 3 Sketch of cut face of thin section (50 μm) of marble for analysing contiguous grains of calcite crystals. The dotted areas were analysed. C and O indicate $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values, respectively. Solid area shows graphite crystal.

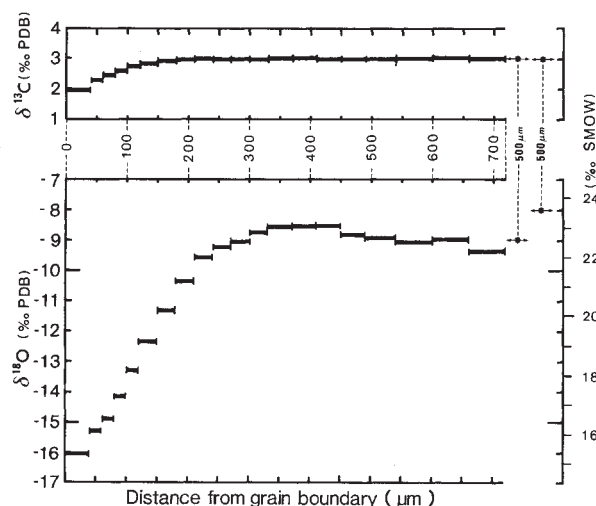


Fig. 4 Isotopic results for calcite crystals of traverse C at the grain boundary contact with calcite crystal in different crystals of the same specimen as shown in Figs 1 and 3. Two data for the core parts at about 500- μm intervals toward the crystal core of traverse C are plotted as asterisks. The cross-sectional area of the calcite prism along the traverse C is about 0.9 mm².

Such isotopic zoning has implications for isotope thermometry. The weighted average of the carbon isotope composition of graphite coincides with that of its inner zone within our analytical error. The weighted average of the carbon isotope composition of calcite in traverse C, ~700 μm from the boundary, also coincides with the levelled value of the core. Therefore, the calculated carbon isotope temperature will not be influenced by retrograde metamorphism, providing similar single graphite crystals are selected. The weighted averages of the oxygen isotope composition of the marginal calcite 200 μm from the boundary of traverse A and C are -6.4 and -12.6‰, respectively. These values are ~1 and 3‰ less than that of the core, respectively. Assuming a spherical calcite 5 mm in diameter, the effective volume of the outer 200 μm is about 22 per cent of the total volume. Therefore, these heterogeneities at the margins of the calcites for traverses A and C will shift the average values about -0.2 and -0.6‰, respectively. This agrees with the remark by Schwarcz *et al.*¹⁵ that quartz-calcite fractionations appear to be affected more by retrograde exchange than silicate-oxide fractionation (for example, quartz-magnetite). As pointed out by Valley¹⁶, sharp gradients in isotopic composition measured in metamorphic rocks represent the time-integrated effect of exchange and it is difficult to resolve the effects of individual geological events. However, isotopic zoning may provide another means to elucidate the retrograde or polymetamorphic processes in nature.

This work was supported in part by the Grant in Aid from The Ministry of Education, Science and Culture of Japan and The Itoh Science Foundation. I thank Dr N. Niitsuma for his advice on microscale isotopic analyses and for discussions, R. Ito and H. Kitagawa for technical assistance and Professors T. Sameshima, C. A. M. Brennkmeijer, J. Hoefs, J. W. Valley, E. L. Grossman and W. Lamb who critically read the manuscript.

Received 8 June; accepted 7 October 1987.

1. Rumble, D. III *Rev. Miner.* **10**, 327-353 (1982).
2. Valley, J. W. & O'Neil, J. R. *Contrib. Miner. Petrol.* **87**, 158-173 (1984).
3. Garlick, G. D. & Epstein, S. *Geochim. cosmochim. Acta* **31**, 181-214 (1967).
4. Valley, J. W. & O'Neil, J. R. *Geochim. cosmochim. Acta* **45**, 411-420 (1981).
5. Wada, H. & Suzuki, K. *Geochim. cosmochim. Acta* **47**, 697-706 (1983).
6. Morikyo, T. *Contrib. Miner. Petrol.* **87**, 251-259 (1984).
7. Kreulen, R. & Van Beek, P. C. J. M. *Geochim. cosmochim. Acta* **47**, 1527-1530 (1983).
8. Arnett, J. D., Schidlowski, M., Sarbos, U., Georg, U. & Amstutz, G. C. *Geochim. cosmochim. Acta* **49**, 1553-1560 (1985).
9. Weis, P. L. *Geology* **8**, 296-297 (1980).

10. Wada, H., Ito, R. & Akiyama, F. *Geosci. Rept. Shizuoka Univ.* **10**, 133-141 (1984).
11. Jones, L. M. *et al. Terra Cognita* **6**, 263 (1986).
12. Eldridge, C. S. *et al. Terra Cognita* **6**, 134 (1986).
13. Bottinga, Y. *Geochim. cosmochim. Acta* **33**, 49-64 (1969).
14. Suzuki, M. *J. Sci., Hiroshima Univ.* **7**, 217-296 (1977).
15. Schwarcz, H. P., Clayton, R. N. & Mayeda, T. K. *Bull. geol. Soc. Am.* **81**, 2299-2316 (1970).
16. Valley, J. W. *Rev. Miner.* **16**, 445-490 (1986).
17. Wada, H. *J. geol. Soc., Japan* **88**, 741-751 (1982).
18. Wada, H., Niitsuma, N. & Saito, T. *Geosci. Rept. Shizuoka Univ.* **7**, 35-50 (1982).
19. Wada, H., Fujii, N. & Niitsuma, N. *Geosci. Rept. Shizuoka Univ.* **10**, 103-112 (1984).

The movement of hydrocarbons in shales

A. S. Mackenzie*, D. Leythaeuser†, P. Muller†‡, T. M. Quigley* & M. Radke†

* BP Petroleum Development Ltd, Britannic House, Moor Lane, London EC2Y 9BU, UK

† KFA-Jülich, Postfach 1913, D-5170 Jülich, FRG

Oil and gas hydrocarbons are formed at temperatures above 100 °C by the chemical breakdown of the organic remains of dead organisms preserved in fine-grained rocks called source rocks. The organic remains are known as kerogen. If enough hydrocarbons are formed, some are expelled from the source rocks. We have analysed Kimmeridge Clay (Upper Jurassic) shales from the North Sea which are currently expelling hydrocarbons into adjacent sandstones. The results could be explained if most expulsion occurs by the pressure-driven flow of a discrete hydrocarbon phase through the pores of shales and into the pores of the sandstones, as suggested previously¹. In addition, capillary forces seem to increase the efficiency of expulsion close to the interfaces between the shales and interbedded sandstones. It appears that molecular diffusion of hydrocarbons with less than about ten carbon atoms alters the composition of the hydrocarbon fluids remaining in, and expelled from, the shales.

Samples of Kimmeridge Clay shales came from the Brae region of the North Sea (~58°45'N, 1°20'E). Here the shales are interbedded with sandstones in which oil derived from the shales has accumulated². Over 350 samples were collected as conventional cores from two wells in the region (Fig. 1). The current temperatures of the sampled horizons are between 100

‡ Present address: Universität Bremen, Postfach 330440, 2800 Bremen 33, FRG.

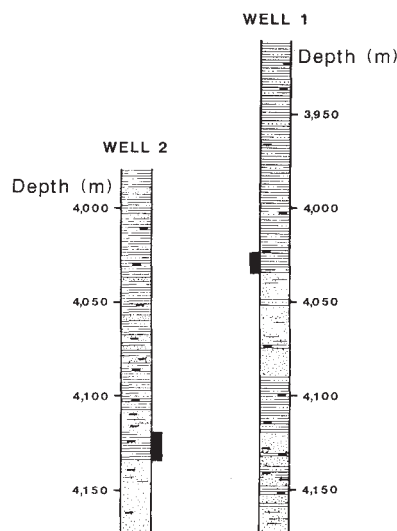


Fig. 1 Geological columns for two wells studied. Lines represent shales and dots sandstones. Solid blocks alongside columns indicate zones selected for detailed study. Depths are measured from the deck of the drilling platform, which was 25 m above sea level.

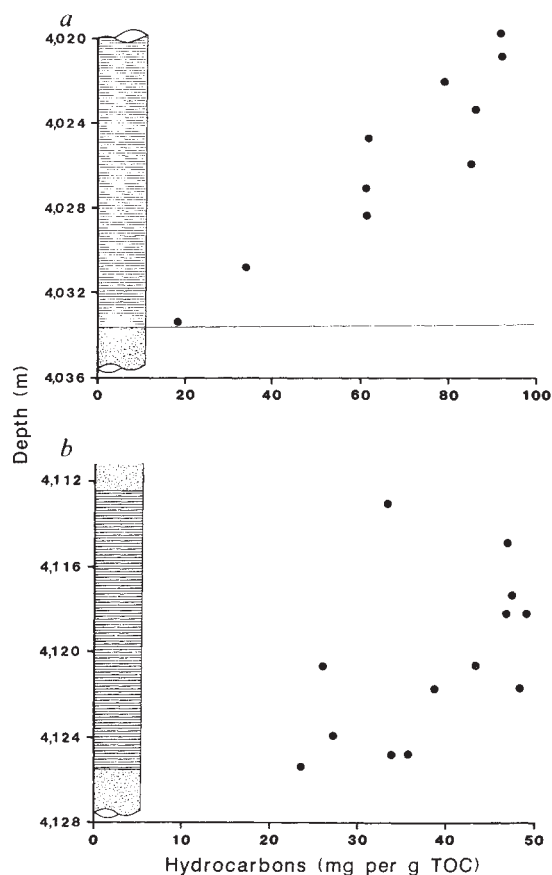


Fig. 2 Yields of C_{15+} hydrocarbons and solvent-extractable (soluble) organic matter relative to total organic carbon as a function of depth for *a*, a 13 m interval at the base of a thick shale from Well 1 and *b*, a 14 m shale from Well 2. The yields were measured by standard methods¹⁰.

and 120 °C (ref. 3), which is the temperature range over which expulsion of oil from most source rocks begins³.

Elsewhere we have demonstrated that a significant amount of the kerogen in the samples of Kimmeridge Clay studied has been converted to hydrocarbons⁴; also that at depths greater than 4 km below sea level, a substantial fraction of the hydrocarbons has been expelled. We also attempted to demonstrate⁴ the likelihood that the initial concentrations and compositions of kerogen and hydrocarbon in the samples studied were similar before major hydrocarbon formation and expulsion began. Therefore any systematic variation in the concentration and composition of hydrocarbons between samples is probably caused by differences in their histories of hydrocarbon formation and expulsion.

Figure 2 plots the mass of hydrocarbons relative to total organic carbon as a function of depth for the lower margin of a thick Kimmeridge Clay interval which overlies a sandstone in Well 1 (see Fig. 1), and of a 14 m-thick Kimmeridge Clay interval interbedded with a sandstone in Well 2 (see Fig. 1). The hydrocarbons decrease in concentration towards the margins of the shales (Fig. 2). This suggests that the movement of hydrocarbons from the centres to the margins of the shales is slower than the movement of hydrocarbons from the margins into the sandstones, as we have already observed⁵.

Figure 3 plots the concentrations of four individual hydrocarbons, namely *n*-pentane (*n*-C₅), *n*-undecane (*n*-C₁₁), *n*-pentadecane (*n*-C₁₅), and *n*-eicosane (*n*-C₂₀), relative to total organic carbon for the same samples as Fig. 2. Although *n*-C₁₁ to *n*-C₂₀ decrease in concentration towards the margins of the shales, the concentration of *n*-C₅ either increases towards the

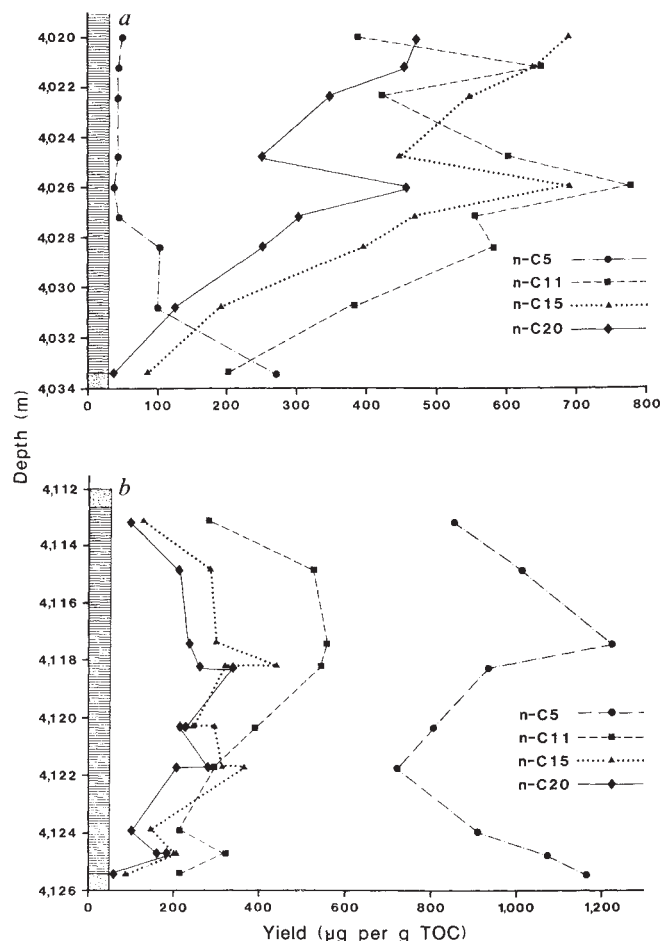


Fig. 3 Yields of *n*-C₅, *n*-C₁₁, *n*-C₁₅ and *n*-C₂₀ alkanes relative to total organic carbon for *a*, a 13 m interval at the base of a thick shale from Well 1 and *b*, a 14 m shale from well 2. The C₁₅ and C₂₀ alkanes were quantified by gas chromatography analysis of the solvent-soluble hydrocarbons, as described elsewhere⁵. The C₅ and C₁₁ alkanes were quantified by thermal vaporization-gas chromatography of dry samples. 60 mg rock powder were rapidly heated to 250 °C in a helium stream and held at that temperature for 9 min. Hydrocarbons liberated by this treatment were trapped on silica gel at the temperature of liquid nitrogen, then analysed by capillary gas chromatography. For quantification, known amounts of an internal standard were added to the rock samples before analysis. The equipment used was a modified DCS-820 GS source rock system.

margin (Fig. 3*a*) or remains roughly constant (Fig. 3*b*). Hence it appears that the larger *n*-C₁₁ to *n*-C₂₀ molecules are more readily expelled than the smaller *n*-C₅ molecules. Because *n*-C₅ is generally more mobile than *n*-C₁₁ to *n*-C₂₀, this observation seems difficult to explain. But the results summarized in Figs 2 and 3 can be rationalized by an expulsion mechanism, whereby most of the hydrocarbons move from the shales to the sandstones by pressure-driven flow of a discrete hydrocarbon phase, and are perturbed by diffusion of hydrocarbons with less than 11 carbon atoms through the organic phases in both the shales and sandstones.

Figure 4*a* and *b* attempts to explain this model by dividing into three stages our perceived evolution of the concentrations of hydrocarbons expressed (relative to total organic matter) in the 14 m shale and at the shale margin. The conversion of kerogen to hydrocarbons occurs over a range of temperatures. Stage 1 represents the situation at low temperatures within this range. Only small amounts of hydrocarbons have been formed, and these are mainly present as isolated droplets of a hydrocarbon phase, which are incapable of flowing out of the source

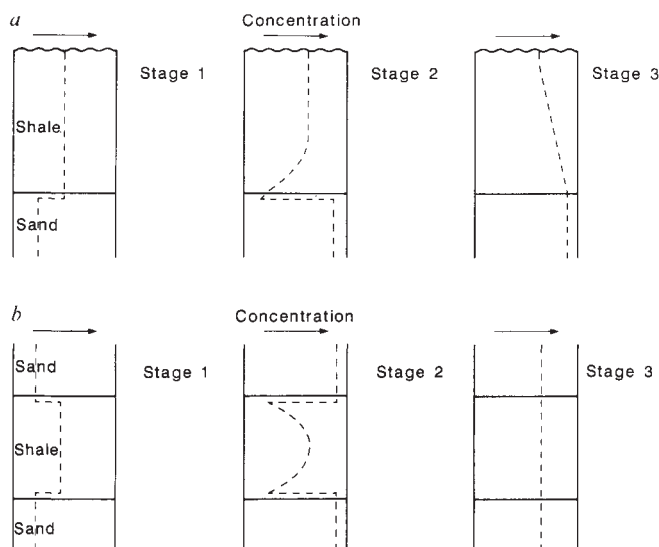


Fig. 4 Schematic evolution of concentrations of individual hydrocarbons relative to total organic matter as a function of time for a, the lower margin of a thick shale and b, a shale band.

rock whose pores are otherwise filled with water. Hence the hydrocarbons are only found in rocks which include significant amounts of kerogen. Because these are mainly restricted to shales in siliciclastic sequences, the hydrocarbon concentrations are high in the shales and very low in the adjacent sandstones. Stage 2 corresponds to the onset of the expulsion of hydrocarbons from shales. The source rock has been raised to higher temperatures than in stage 1 (this is normally achieved by burial to greater depths), and sufficient hydrocarbons have been formed to create an interconnected hydrocarbon phase alongside the aqueous phase in the pores of the shales. We have shown that, in many cases at depths greater than 2.5 km below sea level, the pore fluids (water and hydrocarbons) of shales are at higher pressures than the pore fluids of interbedded sandstones⁶. In such cases, interconnected hydrocarbon stringers bleed from the pores of the shales into the pores of the sandstones. We initially thought that the greater expulsion of hydrocarbons could be explained by a decrease in porosity towards the margin; subsequent measurement of porosity showed that this is not the case⁴ and that an alternative explanation is required.

If the walls of some of the pores of the shales and sandstones are water-wet then, according to the laws of capillary pressure⁷, the pressure in the hydrocarbon phase in such pores exceeds that in the aqueous phase. The pressure contrast between hydrocarbon and water phases increases with decreasing pore size⁷. Therefore at the interfaces between water-wet pores in small-pored shales and large-pored sandstones, there is an additional pressure difference between the hydrocarbon phases in the two media. This increases the flow of hydrocarbons from shales to sandstones. Although it could be argued that the pores of organic-rich rocks are entirely oil-wet, our results (Fig. 2a and b) do not support this: the decreased concentrations towards the shale margins imply that a significant fraction of the pores

are water-wet. Indeed, most experiments suggest that silicate minerals are preferentially water-wetting⁸.

The types of hydrocarbon-concentration profiles which we envisage developing during stage 2 of the model are shown in Fig. 4. In the shales, the lowest concentrations occur close to the interfaces with the sandstones because of the effects of capillary forces. We have expressed the hydrocarbon concentrations relative to total organic matter. In the shales organic matter comprises chiefly hydrocarbons and kerogen, but in the sandstones the organic matter is almost entirely hydrocarbons derived by expulsion from the shales. Therefore during stage 2 the concentration of hydrocarbons in the total organic matter is greater in the sandstones than the shales. This is reflected in Fig. 4.

So during stage 2, strong gradients in the concentration of hydrocarbons in total organic matter evolve. If a hydrocarbon is strongly diffusive it can move through the organic phases by random motion to reduce these gradients to the concentration profiles shown as stage 3 in Fig. 4. If the hydrocarbons are weakly diffusive no such effect is possible, and stage 2 concentration profiles are preserved. Thus, the margins of the shales would be highly depleted in weakly diffusive hydrocarbons, and enriched in strongly diffusive hydrocarbons. Inspection of Fig. 3 suggests $n\text{-C}_5$ should be classified as strongly diffusive, whereas $n\text{-C}_{11}$, $n\text{-C}_{15}$ and $n\text{-C}_{20}$ are weakly diffusive. We have modelled this effect and obtain reasonable results if the diffusion coefficients are about $10^{-13} \text{ m}^2 \text{ s}^{-1}$ for $n\text{-C}_5$ and less than $10^{-15} \text{ m}^2 \text{ s}^{-1}$ for $n\text{-C}_{11}$, $n\text{-C}_{15}$ and $n\text{-C}_{20}$ (ref. 3).

It is important to recognize that because we express our hydrocarbon concentrations relative to total organic matter, the diffusion we invoke must occur through all, or part of, the non-hydrocarbon organic phases (such as kerogen), as well as the hydrocarbon phases. If the hydrocarbon concentrations were expressed relative to total rock weight, or relative to total hydrocarbons (implying random diffusion through the whole rock, or restricted to the hydrocarbon phases), large concentration gradients would not develop between shales and sandstones during stage 2; also, the driving force for the diffusion of C_5 hydrocarbons from the sandstone to the shale, as suggested in Fig. 3, would be much less.

In other shale-source rock/sand sequences, we also observed major fractionation of hydrocarbons with more than 10 carbon atoms (for example, $n\text{-C}_{15}$ was expelled more readily than $n\text{-C}_{20}$) (ref. 5). We believe that in these sequences the hydrocarbon mixtures formed in the source rock were richer in $\text{C}_1\text{-C}_5$ hydrocarbons than in the case described here. Perhaps at the pressure-temperature conditions during expulsion, the hydrocarbons of the source rocks studied previously⁵ were present as two phases, namely hydrocarbon gas and hydrocarbon liquid (or oil). The lower viscosity gas was expelled preferentially by the processes described earlier, and the hydrocarbons could only escape by dissolution in the gas. Solubility increases with decreasing density⁹: $n\text{-C}_{15}$ is more soluble in gas than $n\text{-C}_{20}$ and therefore is expelled more readily. Perhaps because the mole fractions of $\text{C}_1\text{-C}_5$ hydrocarbons are lower in the Brae example studied here, a separate gas phase has not evolved; all the hydrocarbons remain single phase and liquid, and hence no fractionation of hydrocarbons with more than 10 carbon atoms occurs.

Received 10 August; accepted 18 November 1987.

1. Tissot, B. P. & Welte, D. H. *Petroleum Formation and Occurrence*. (Springer, Berlin, 1978).
2. Harms, J. C., Talkenber, P., Pickles, E. & Pollock, R. E. in *Petroleum Geology of the Continental Shelf of Northwest Europe* (eds Illing, L. V. & Hobson, G. D.) 380-391 (Heyden, London, 1981).
3. Coates, G. P., Mackenzie, A. S., Quigley, T. M. *Org. Geochem.* **10**, 235-245 (1986).
4. Mackenzie, A. S., Price, I., Leythaeuser, D., Müller, P., Radke, M. & Schaefer, R. G. in *Petroleum Geology of the Continental Shelf of Northwest Europe* (ed. Glennie, K. W.) (Heyden, in the press).

5. Leythaeuser, D., Mackenzie, A. S., Schaefer, R. G. & Bjørøy, M. *Bull. Am. Ass. Petrol. Geol.* **68**, 196-219 (1984).
6. England, W. E., Mackenzie, A. S., Mann, D. M. & Quigley, T. M. *J. geol. Soc. Lond.* **144**, 327-347 (1987).
7. Hinch, E. J. *J. Phys. Chem. Hydrodynam.* **6**, 601-622 (1985).
8. Berner, F. C. & Bartell, F. E. *Drill Prod. Pract.* **341**, 209-219 (1941).
9. Price, L. C., Wegner, L. M., Ging, T., Blout, G. W. *Org. Geochem.* **4**, 201-221 (1981).
10. Radke, M., Sittardt, H. G. & Welte, D. H. *Analyt. Chem.* **50**, 663-665 (1978).

A new form of reproductive parasitism in cliff swallows

Charles R. Brown & Mary Bomberger Brown

Department of Biology, Yale University, PO Box 6666, New Haven, Connecticut 06511, USA

A common reproductive strategy among some egg-laying animals, especially birds, is to lay an egg in another individual's nest and thereby parasitize the reproductive effort of others of either the same species or a different species. Intraspecific parasitism is now known to occur regularly in some species¹⁻⁴ and sporadically in many others^{5,6}, and may represent a strategy by which individuals augment their reproductive performance² or succeed in reproducing when it would otherwise be impossible or too costly⁵. We have discovered that colonial cliff swallows (*Hirundo pyrrhonota*) not only lay eggs in other individuals' nests, but also physically transfer eggs between nests after the eggs are laid. Egg transfers can occur at any time after an egg is laid and before it hatches, and may represent a strategy by which an individual distributes its eggs in several nests to ensure some surviving offspring in the event of nesting failures. Sneaky transfer of eggs between nests represents a previously unknown form of reproductive parasitism in birds.

Cliff swallows build gourd-shaped nests out of mud pellets which are fastened in dense colonies underneath overhanging rock ledges on the sides of cliffs and canyons and, more recently, underneath bridges and in highway culverts throughout much of western North America. Breeding within a colony is highly synchronous and colonies allow cliff swallows to gain information from each other on the whereabouts of flying insect prey⁷. Females parasitize other individuals' nests by sneaking into nests momentarily left unattended and laying eggs there². At our study site near Ogallala in Keith County, Nebraska, cliff swallows arrive in early May each year and remain until mid August, when they migrate to their wintering range in southern South America⁸. In Nebraska these birds nest both solitarily and in colonies ranging from 2 to 3,500 nests in size (mean colony size = 355 nests, s.d. = 561, $n = 276$ colonies).

While studying cliff swallows from 1982-1987, three sorts of evidence for transfers of eggs between nests were found: (1) direct observations of birds transferring eggs, (2) movement of marked eggs between nests, and (3) appearance in nests, after incubation began, of eggs that hatched at the same time as the rest of the clutch.

Upon their arrival in the spring, cliff swallows at selected colonies were captured in mist nets and their white forehead patches were painted in unique colour combinations for individual recognition². Between 50 and 80% of the nest owners in samples of 45-75 nests in each colony were individually marked. We watched nests continuously for ~75% of the daylight hours, beginning before egg laying and continuing during part of incubation at each colony. Samples of nests in six colonies that ranged in total size from 125 to 1,100 nests were observed.

We observed two definite and three probable instances of cliff swallows carrying eggs to other nests. In two instances a bird left its own nest with an intact egg between its mandibles and flew to a neighbouring nest. In one case the owner of the neighbouring nest was not present and the intruder entered, deposited the egg in the nest, and returned to its own nest. In the second case one owner of the neighbouring nest was present; a fight ensued when the intruder entered with the egg. The intruder was evicted from the nest within 10 s, but the egg remained in the nest. In the three remaining instances, a cliff swallow emerged from its own nest carrying an egg between its mandibles, but we lost sight of the bird when it flew towards the opposite end of the colony. We are certain that these birds did not drop the eggs because the entire colony was located

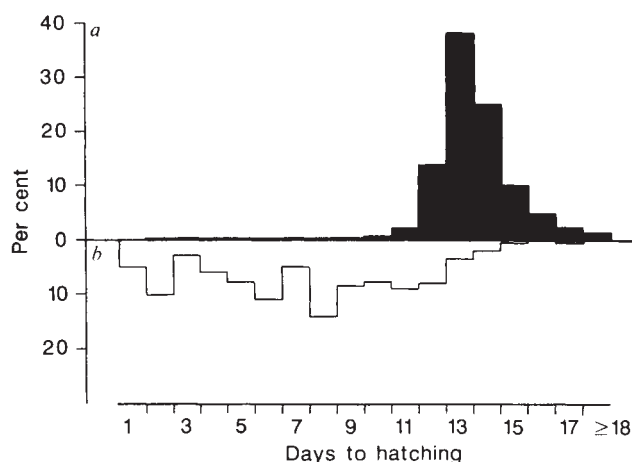


Fig. 1 *a*, Percentage occurrence of time periods from the laying of the last egg until the hatching of the first egg for whole cliff swallow clutches ($n = 3,370$). *b*, Percentage occurrence of time periods from appearance of an egg added to a clutch by transfer, until hatching of the first egg in the clutch ($n = 357$).

over water and we did not see or hear any splashes. In each case the bird either transferred the egg to another nest unseen by us, or carried the egg away from the colony. Three of the five perpetrators of transfers were females, and the sex of the remaining two individuals was unknown. The five transfers directly observed occurred in colonies of 750 and 1,100 nests, the two largest colonies which were intensively watched.

We also recorded the transfer of numbered eggs. Each nest's number was written in several places on the outside of each egg in a clutch with a Sharpie magic marker. All clutches marked were in stages of incubation and appeared to be complete. We marked a total 204 eggs from 50 nests in three colonies of 90, 120 and 1,100 nests. After marking we returned every second day and checked the contents of each marked nest and all neighbouring unmarked nests. Within two days of numbering, three of the 50 nests (6%) had acquired an intact numbered egg from a neighbouring nest. The clutches from which these transferred eggs came had decreased by one marked egg in two cases and by two marked eggs in the other.

An indirect measure of the frequency of egg transfers is how often eggs appear in nests during incubation yet still hatch in synchrony with the clutch to which they were added. As all cliff-swallow eggs presumably require a reasonably constant period of incubation (12-14 days), any parasitic egg that appears after a host starts incubation, yet still hatches with the host's own eggs, must have been incubated elsewhere for some period of time before being transferred. We used appearance of eggs in a nest three or more days after the clutch size there had stopped increasing and incubation had presumably started, as evidence of a transfer. These eggs were in fact likely to have been physically transferred, because in 25 actual observed cases of parasitic egg laying, no eggs were laid in a host's nest more than two days after the host had completed laying (unpublished data).

The contents of 5,077 nests in 46 colonies were checked every (or every second) day during this study, beginning before eggs were laid and continuing until clutches hatched. One or more transferred eggs appeared in 306 of 4,821 nests (6.3%). This percentage occurrence of transferred eggs agreed closely with that obtained by marking eggs (6%). The time elapsing between the arrival of a transferred egg in a nest and the start of hatching in that nest varied between 1 and 17 days (mean = 7.33, s.d. = 3.58), a considerably shorter interval than the normal clutch incubation period (mean = 13.58, s.d. = 1.85) (Fig. 1). These 306 nests had 384 eggs transferred to them; 70 nests had multiple transfers (2-5 eggs) occurring either simultaneously or sequen-

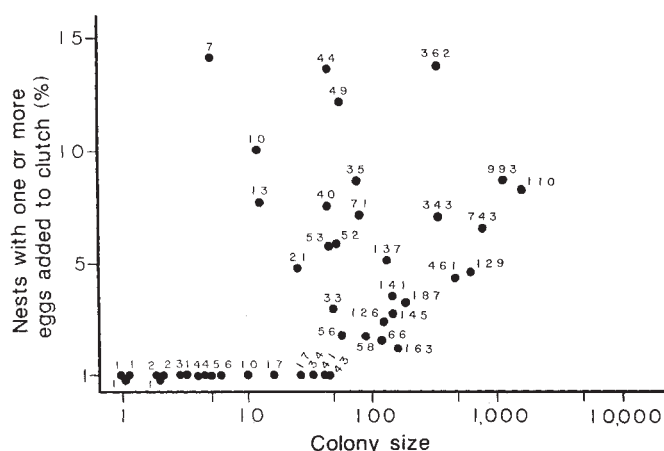


Fig. 2 Percentage nests containing one or more eggs added to a clutch by transfer versus cliff swallow colony size (number of active nests). Sample size is shown for each colony. When all colonies were considered, percentage nests with one or more eggs added increased significantly with colony size ($r_s = 0.60$, $P < 0.001$). When only colonies with more than 10 nests were considered, there was no significant correlation between percentage nests with one or more eggs added and colony size (Spearman rank correlation coefficient, $r_s = 0.21$, $P = 0.25$).

tially. Of the 384 transferred eggs, 42 (10.9%) definitely did not produce fledged offspring because the nests containing them failed; 166 eggs (43.2%) hatched in synchrony with the host's clutch and definitely produced fledged offspring; and the fate of 176 eggs (45.9%) was unclear. Among the transferred eggs detected by direct observation and egg numbering ($n = 5$), one egg definitely produced offspring, while the fate of the remaining four was uncertain.

Of the 273 nests containing transferred eggs which produced some nestlings, in 271 (99.3%) all viable eggs including the transfers hatched within 24–36 h of each other. The remaining two nests contained an egg (probably the transfer) that hatched 10 days after the others. Thus, the interval between the time a transferred egg appeared in a nest and hatching (Fig. 1) indicated how long the egg had been incubated elsewhere before being transferred. Cliff swallows transferred eggs to other nests virtually any time after laying, even one and two days before an egg was due to hatch (Fig. 1). The birds directly seen to transfer eggs were still laying in their own nests and had not begun incubation. A small percentage of entire clutches (Fig. 1a), had shorter than expected incubation periods. These clutches did not meet our criterion for classification as egg transfers because all the eggs appeared in a normal sequence. Because of the short incubation periods, however, they probably did represent egg transfers, perhaps transfers of entire clutches.

A large cliff-swallow colony potentially might contain more suitable nests to which to transfer an egg, and therefore the incidence of transfers might increase with colony size. Colony size affected the incidence of nests with at least one transferred egg (Fig. 2), but only when small colonies, ≤ 10 nests in size, were included. In 14 colonies with 10 or fewer nests, only one transfer was detected ($n = 48$ nests). When only colonies with more than 10 nests were considered, colony size did not affect percentage occurrence of transferred eggs (Fig. 2). This result is perhaps not surprising as many of the interactions among cliff swallows within a colony involve only close neighbours^{2,7}. Of the five cases in which the perpetrator of the transfer was known and the nest to which it transferred an egg was also known, three occurred between nests that were adjacent to each other (15 cm apart) in the colony. In the other two cases, five nests (52 cm) and 22 nests (112 cm), respectively, separated the perpetrators' and hosts' nests.

Transfer of eggs may be a sophisticated behavioural strategy involving subtle assessment of potential host individuals by

potential transferers and removal of some of the host's eggs in advance. Of the 306 nests known to have an egg transferred to them, 33 failed (10.8%). This is less than half of the total nest failure rate for cliff-swallow nests in our study population as a whole (1,102 of 4,708 nests failed, 23.4%), suggesting that transferers may select superior neighbours as hosts. Potential transferers might remove a host's egg in advance of adding one, as is known for other intraspecific (ref. 9; H. W. Power, personal communication) and some interspecific^{10,11} parasites and for cliff swallows when laying eggs in hosts' nests (unpublished data). This is suggested by the fact that disappearance of a single egg from the host's nest occurred within the 1–4 day period immediately preceding the appearance of the transferred egg, for 125 of the 377 transferred eggs (33.2%) for which past nest histories were known. Instances of single eggs disappearing from clutches are often caused by intruding cliff swallows that toss out eggs^{2,12}. The percentage of host nests suffering single-egg losses before a transfer (33.2%) is over three times that of cliff-swallow nests in our study population as a whole (479 of 4,899 nests with single-egg losses, 9.8%).

How might an individual benefit by transferring eggs? As many transfers occur after the perpetrator has ceased laying (Fig. 1), reproductive output probably cannot be supplemented by transferring eggs. After incubation starts an individual cannot lay more eggs in its own nest to replace those transferred elsewhere. Instead, transfer of eggs might increase the chances of fledging at least some offspring in a risky environment. In their ancestral nesting habitat—rocky cliffs and canyons—cliff swallows are often affected by inclement weather and rock slides which can destroy many nests (unpublished data). Spreading a clutch of eggs around more than one nest could insure against nesting failure^{13,14}.

There may be some cost to transferring eggs. In seven of the eight cases in which the identity of the transferer was known, a parasitic egg had been previously added to the transferer's own nest either by laying or transfer. This suggests that in assessing which nearby nests are candidates for an egg transfer, a transferer may leave its own nest unattended to the degree that it is more likely to be parasitized itself. Transferers are probably not simply removing someone else's parasitic eggs from their nest, as these birds are unable to recognize eggs².

Our estimate that about 6% of cliff-swallow nests contain transferred eggs is undoubtedly an underestimate. Eggs that are transferred during laying would resemble, in nest-check data, parasitic eggs laid in a nest and go undetected using our criterion (Fig. 1). Colour-marked birds did in fact transfer some eggs during laying. Estimating the true frequency of egg transfers and parasitic egg laying in cliff-swallow colonies may have to await the development of DNA-fingerprinting techniques for precise assignment of parentage. Incredible though egg transfer in cliff swallows may seem, it has in fact been reported (rarely) in woodpeckers^{15,16} and in corvids¹⁷ and may prove to be more common in birds than has previously been supposed.

We thank Cathy Boersma, Carol Brashears, Karen Brown, Rachel Budelsky, Laurie Doss, Jerri Hoskyn, Laura Jackson, Deborah Johnson, Kathi Miller, Todd Scarlett, Martin Shaffer, Lora Sherman for field assistance, John Janovy, Jr and Anthony Joern for use of the facilities of the Cedar Point Biological Station, and Harry Power and Bridget Stutchbury for helpful comments. This work was supported by the NSF, the National Geographic Society, Princeton and Yale universities, the Chapman Fund of the American Museum of Natural History, the Bache Fund of the National Academy of Sciences, Sigma Xi, Alpha Chi and Raymond and Kathryn Brown.

Received 7 October; accepted 13 November 1987.

- Andersson, M. & Eriksson, M. O. G. *Am. Nat.* **120**, 1–16 (1982).
- Brown, C. R. *Science* **224**, 518–519 (1984).
- Emlen, S. T. & Wrege, P. H. *Ethology* **71**, 2–29 (1986).
- Power, H. W., Litovich, E. & Lombardo, M. P. *Auk* **98**, 386–388 (1981).

5. Andersson, M. in *Producers and Scroungers: Strategies of Exploitation and Parasitism* (ed. Barnard, C. J.) 195–228 (Croom Helm, London, 1984).
6. Yom-Tov, Y. *Biol. Rev.* **55**, 93–108 (1980).
7. Brown, C. R. *Science* **234**, 83–85 (1986).
8. American Ornithologist's Union, *Check-list of North American Birds*, 6th edn (Lawrence, Kansas, 1983).
9. Feare, C. *The Starling* (Oxford University Press, Oxford, 1984).
10. Friedmann, H. *The Cowbirds* (Thomas Springfield, Illinois, 1929).
11. Hamilton, W. J. III & Orians, G. H. *Condor* **67**, 361–382 (1965).
12. Brown, C. R. thesis, Univ. Princeton, (1985).
13. Gillespie, J. H. *Genetics* **76**, 601–606 (1974).
14. Payne, R. B. A. *Rev. Ecol. Syst.* **8**, 1–28 (1977).
15. Blomme, C. *Ontario Field Biol.* **37**, 34–35 (1983).
16. Truslow, F. K. *Living Bird* **6**, 227–236 (1967).
17. Trost, C. H. & Webb, C. L. *Anim. Behav.* **34**, 294–295 (1986).

Neuronal correlate of pictorial short-term memory in the primate temporal cortex

Yasushi Miyashita & Han Soo Chang

Department of Physiology, Faculty of Medicine, University of Tokyo, Hongo, Tokyo 113, Japan

It has been proposed that visual-memory traces are located in the temporal lobes of the cerebral cortex, as electric stimulation of this area in humans results in recall of imagery¹. Lesions in this area also affect recognition of an object after a delay in both humans^{2,3} and monkeys^{4–7}, indicating a role in short-term memory of images⁸. Single-unit recordings from the temporal cortex have shown that some neurons continue to fire when one of two or four colours are to be remembered temporarily⁹. But neuronal responses selective to specific complex objects^{10–18}, including hands^{10,13} and faces^{13,16,17}, cease soon after the offset of stimulus presentation^{10–18}. These results led to the question of whether any of these neurons could serve the memory of complex objects. We report here a group of shape-selective neurons in an anterior ventral part of the temporal cortex of monkeys that exhibited sustained activity during the delay period of a visual short-term memory task. The activity was highly selective for the pictorial information to be memorized and was independent of the physical attributes such as size, orientation, colour or position of the object. These observations show that the delay activity represents the short-term memory of the categorized percept of a picture.

In a trial of our visual short-term memory task, sample and match stimuli were successively presented on a video monitor, each for 0.2 s at a 16 s delay interval (Fig. 2a). The stimuli were newly selected for each trial among 100 computer-generated coloured fractal patterns (examples are shown in Fig. 1) and 100 pseudo-coloured images of scenery. Two monkeys (*Macaca fuscata*) were trained to memorize the sample stimulus and to decide whether the match stimulus was the same or different (see legend to Fig. 2). Extracellular spike discharges of 188 neurons were recorded from the anterior ventral part of the temporal cortex (Fig. 2b) of these monkeys with standard physiological techniques¹⁹. Recording of electro-oculographs revealed no systematic differences in eye position which could be related to differential neural responses described below.

Figure 2c shows reproducible stimulus-dependent discharges during the delay obtained in one cell for four different sample stimuli (Fig. 1a–d) (prominent in a and b, but virtually ineffective in c and d). A time course of the delay activity in Fig. 2ci is shown in Fig. 2d, as contrasted with those for six other ineffective stimuli (Fig. 1c–h). These histograms are representative of those accumulated with 57 other sample stimuli. Only two of the 64 tested stimuli (Fig. 2e) were followed by especially high delay activity (>10 impulses $\cdot$ s⁻¹), shown in Fig. 2ci and ii.

These delay activities do not represent mere sensory after-discharge^{9,11} for the following reasons. First, the high rate of firing did not decline throughout the whole 16 s delay period

(Fig. 2d). Second, firing frequency exhibited during the delay was not necessarily correlated with that during the stimulus presentation (Fig. 2c and d). Third, the delay activity in some neurons started after a latency of a few seconds following stimulus presentation (data not shown). Thus, it is concluded that the delay activity is not a passive continuation of the firing during the sensory stimulation, but represents a mnemonic activity to retain visual information.

Of the 188 neurons tested, 144 showed a correlation between firing and one or more events of the trials. Among the 144 cells, 95 showed a sustained increase or decrease of discharge frequency during the delay period, whereas the others fired only during stimulus presentation. In 77 of these 95 cells, the discharge frequency varied depending on sample stimuli, but the remaining 18 did not exhibit such selectivity. In many of the 77 selective cells, only a few pictures elicited a strong delay activation such as shown in Fig. 2c and d. It is notable that the optimal picture differed from cell to cell, and that the whole population of the optimal pictures for the 77 cells covered a substantial part of the repertory of 200 pictorial stimuli.

For further analysis of triggering features of the delay responses, sample pictures were manipulated in the following way (for example Fig. 1i–l): (1) stimulus size was reduced by half, (2) stimuli were rotated by 90° in a clockwise direction, (3) coloured stimuli were transformed into monochrome by referring to a pseudo-colour look-up table, and (4) stimulus position was changed on the video monitor (data not shown) (a 0.2 s stimulus presentation time is short enough to exclude the contribution of saccadic eye movement). Figure 3 shows responses of a neuron which consistently fired during the delay after one particular picture (shown in Fig. 1i–l) but not after others, irrespective of stimulus size (Fig. 3aii and bii), orientation (Fig. 3aiii and biii), or colour (Fig. 3aiv and biv). Similar tolerance of responses was observed in a majority of the tested delay neurons: to size in 16 out of 19 cells, to orientation in 5 out of 7 cells, to colour-monochrome in 15 out of 20 cells, and to position in 8 out of 13 cells. In other neurons, manipulation of the most effective stimulus reduced or abolished the delay discharge.

In the inferior temporal cortex, shape-selective neuronal discharges have been reported for Fourier descriptors¹², face^{13,16,17}, hands^{10,13} or stimuli used in a discrimination task^{14,15}, although all of these were sensory responses evoked during presentation of stimulus. The relative selectivity of these sensory neurons remained invariant over changes of size, position, orientation or contrast^{12–15,17}. The present delay responses could be derived through such sensory responses, inheriting from them the tolerance to such stimulus transformations. It is notable that, for the Fourier descriptor neurons¹², the absolute level of the response varied widely over changes in stimulus size, although this was not the case for many of the present neurons. This may suggest that the present neurons represent more abstract properties of objects (like shape percept) than do such sensory neurons.

In the colour-selective delay neurons previously described⁹, time courses of sustained delay discharge were similar to those found in the present shape-selective delay neurons (compare Fig. 2d or Fig. 3a with Fig. 7 or Fig. 6 of ref. 9). The differential delay activity in the colour task was mainly found in the cortex of the lower bank of the superior temporal sulcus⁹, lying more posteriorly and dorsally to the presently explored area, and the colour cells seemed to be scattered⁹, whereas the present cells tended to group in smaller areas.

A majority of our shape-selective delay neurons were recorded in TE_{av}²⁰ (or TE₁–TE₂²¹) and some in TG_v²⁰ and in area 35. These areas are anatomically designated as the last link from the visual system to limbic memory systems^{4,20,22,23}. Neurons in these areas were visually responsive with a large receptive field²⁴. Impairment of the recognition memory task resulted from lesions including these areas^{4–7}, consistent with the presently-postulated mnemonic role of neurons in these areas.

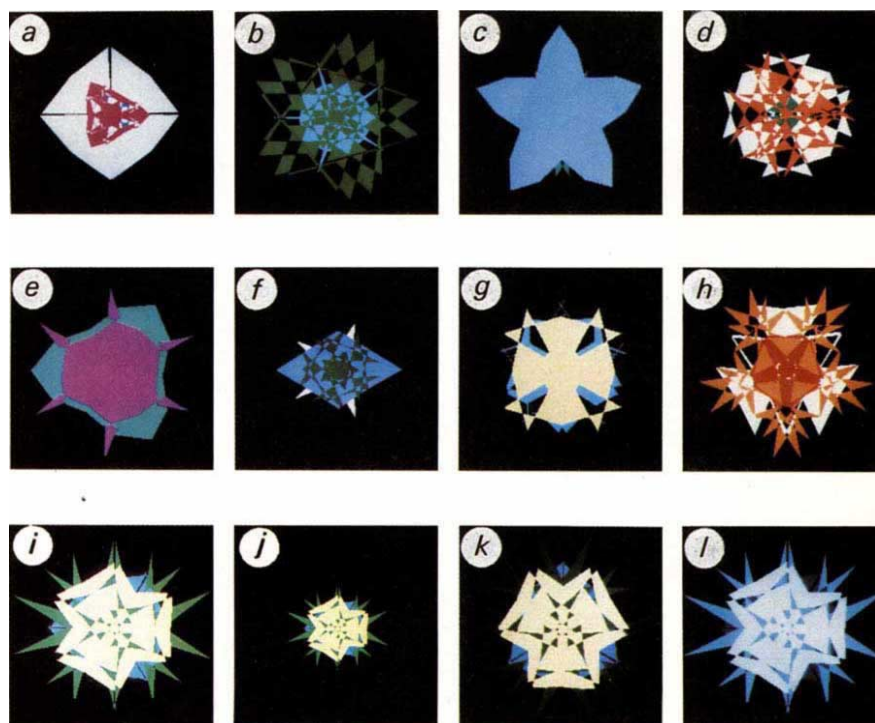


Fig. 1 Examples of coloured fractal patterns. *a-h*, Stimuli used in the trials of Figs 2*c* and *d*. Panels *i-l*, illustrate size reduction (*j*), rotation (*k*) and colour-monochrome transformation (*l*) of stimulus (*i*), as used in the test shown in Fig. 3.

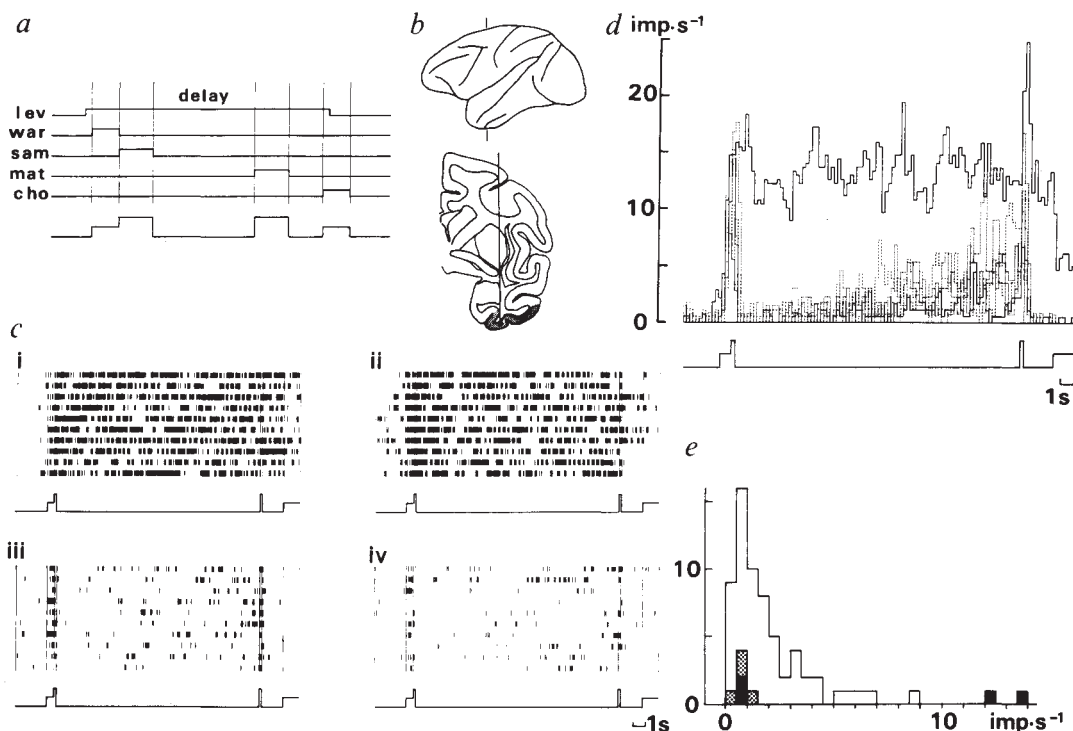


Fig. 2 Responses of a neuron in the anterior ventral temporal cortex in a visual short-term memory task. *a*, Sequence of events in a trial (Lev, lever press by the monkey; War, warning green image (0.5 s); Sam, sample stimulus (0.2 s); Mat, match stimulus (0.2 s) following a delay of 16 s; Cho, choice signal of white image). Lowest trace, the events-chart used in Fig. 2*c-d* and Fig. 3*a*. *b*, Location of recording sites. Top, a lateral view of a macaque brain. Bottom, a section indicated by a vertical line on the lateral view. The stippled area represents the range of recording sites. The vertical line and the dot at its end represent the microelectrode track and the position of the cell shown in Fig. 2*c-e*. *c*, Raster recordings of impulse discharge from a cell. *i*, Trials whose sample stimulus was as shown in Fig. 1*a*. *ii*, *iii* and *iv*, Obtained as in Fig. 1*b*, *c* and *d*, respectively. Trials of the same sample stimulus were originally separated by intervening trials of other sample stimuli, and these were sorted and collected by off-line computation. *d*, Spike-density histograms for the delay activity of Fig. 2*ci* and for those for six other ineffective stimuli (Fig. 1*c-h*). Seven to 16 trials were accumulated for each sample stimulus. Bin width, 200 ms. *e*, Distribution of average delay spike frequencies following 64 different sample pictures, with which more than four trials were tested. The same cell as in *c* and *d*. Ordinate, number of sample pictures used as stimuli. Closed columns, responses shown in Fig. 2*c*. The stippled columns, other responses included in Fig. 2*d*.

Method. Each sample picture was paired with a match stimulus that was identical and one that was not identical. For each trial, the identical or non-identical match stimulus was assigned randomly. If the match stimulus was different from the sample, the monkey had to release the lever and touch the video screen to obtain fruit juice. If the match stimulus was the same as the sample, the monkey had to keep pressing the lever until the choice signal was turned off. If the monkey released the lever before the choice signal, the trial was cancelled. The monkeys' decisions were 85–100% correct. Error trials were excluded from the analysis presented here. In training sessions and the search period stimuli were presented in a fixed sequence. In the recording session, a long sequence of trials using the entire repertory of stimuli was run repeatedly. When some pictures were found to elicit stronger responses than others, the relevant pictures were selected and shorter sequences run with fewer stimuli.

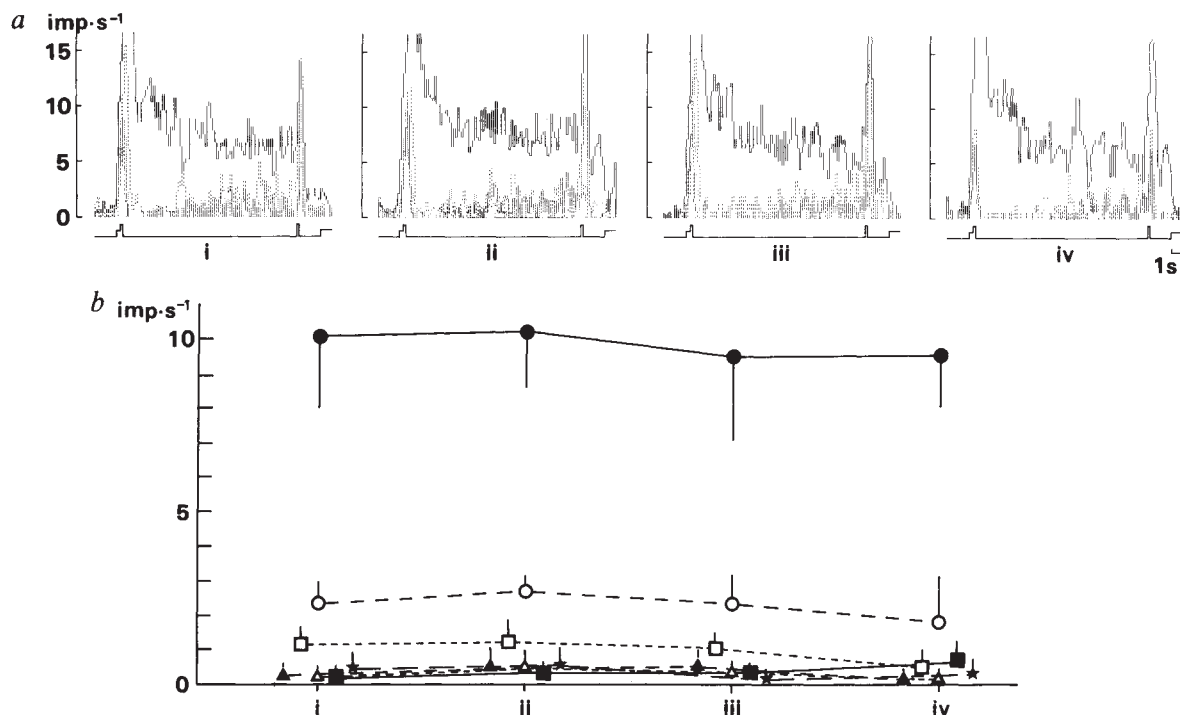


Fig. 3 Response invariance under stimulus transformation in size, orientation or colour. A different cell from that in Fig. 2. *a*, Histograms similar to that in Fig. 2*d*, but for five different sample pictures. *i*, Control responses. Note that the sample stimulus of Fig. 1*i* elicited the strongest delay activity in this cell. *ii*, *iii* and *iv*, Effects of stimulus size reduction by half (for example Fig. 1*j*), of stimulus rotation by 90° in a clockwise direction (for example Fig. 1*k*) and of colour-to-monochrome transformation (for example Fig. 1*l*). *b*, Average delay spike frequencies as a function of stimulus transformation (*i*, original; *ii*, size reduction; *iii*, rotation; *iv*, colour to monochrome). Responses to seven different sample pictures (including five shown in *a*) are plotted with different symbols. ●, Responses to stimuli of Fig. 1*i*–*l*. Error bars indicate standard deviations for 4–15 trials.

The present results suggest that pictorial short-term memory is coded by temporary activation of an ensemble of neurons in the region of the association cortex that processes visual information^{4,9,25}, rather than by neuronal activity in a brain area specialized for short-term memory. Although each neuron in the ensemble has highly abstract and selective coding features, representation of the memory of a picture seems to be distributed among a number of neurons. We need to know how the distributed information is decoded for subsequent decision processes²⁶.

We thank Professor Masao Ito for his encouragement.

Received 3 July; accepted 5 November 1987.

1. Penfield, W. & Perot, P. *Brain* **86**, 595–697 (1963).
2. Kimura, D. *Arch. Neurol.* **8**, 48–55 (1963).
3. Milner, B. *Neuropsychologia* **6**, 191–209 (1968).
4. Mishkin, M. *Phil. Trans. R. Soc. B* **298**, 85–95 (1982).
5. Gaffan, D. & Weiskrantz, L. *Brain Res.* **196**, 373–386 (1980).
6. Sahgal, A., Hutchinson, R., Hughes, R. P. & Iverson, S. D. *Behav. Brain Res.* **8**, 361–373 (1983).
7. Fuster, J. M., Bauer, R. H. & Jervey, J. P. *Expl. Neurol.* **71**, 398–409 (1981).
8. Warrington, E. K. & Shallice, T. *Quart. J. exp. Psychol.* **24**, 30–40 (1972).
9. Fuster, J. M. & Jervey, J. P. *J. Neurosci.* **3**, 361–375 (1982).
10. Gross, C. G., Rocha-Miranda, C. E. & Bender, D. B. *J. Neurophysiol.* **35**, 96–111 (1972).
11. Gross, C. G., Bender, D. B. & Mishkin, M. *Brain Res.* **131**, 227–239 (1977).
12. Schwartz, E. L., Desimone, R., Albright, T. D. & Gross, C. G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5776–5778 (1983).
13. Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. J. *Neurosci.* **4**, 2051–2062 (1984).
14. Sato, T., Kawamura, T. & Iwai, E. *Expl. Brain Res.* **38**, 313–319 (1980).
15. Iwai, E. *Vision Res.* **25**, 425–439 (1985).
16. Perret, D. I., Rolls, E. T. & Caan, W. *Expl. Brain Res.* **47**, 329–342 (1982).
17. Rolls, E. T. & Baylis, G. C. *Brain Res.* **65**, 38–48 (1986).
18. Baylis, G. C. & Rolls, E. T. *Expl. Brain Res.* **65**, 614–622 (1987).
19. Miyashita, Y. & Nagao, S. *J. Physiol., Lond.* **351**, 251–262 (1984).
20. Turner, B. H., Mishkin, M. & Knapp, M. J. *comp. Neurol.* **191**, 515–543 (1980).
21. Seltzer, B. & Pandya, D. N. *Brain Res.* **149**, 1–24 (1978).
22. Herzog, A. G. & Van Hoesen, G. W. *Brain Res.* **115**, 57–69 (1976).
23. Van Hoesen, G. W. & Pandya, D. N. *Brain Res.* **95**, 1–24 (1975).
24. Desimone, R. & Gross, C. G. *Brain Res.* **178**, 363–380 (1979).
25. Anderson, J. R. *Cognitive Psychology and its Implications* (Freeman, San Francisco, 1980).
26. Coltheart, M. *Phil. Trans. R. Soc. B* **302**, 283–294 (1983).

A complete culture system for the chick embryo

M. M. Perry

Gene Expression Group, AFRC Institute of Animal Physiology and Genetics Research, Edinburgh Research Station, Roslin, Midlothian EH25 9PS, UK

The embryonic lifespan of the chick is 22 days. Development in the first day takes place in the oviduct, and in the remaining 21 days in the shelled egg. There have been few attempts to culture oviductal embryos^{1–3}, though methods covering the first few days of development *in ovo* are well established⁴ and a method for the final 18 days of development through hatching has recently been devised⁵. I have now succeeded in culturing the fertilized ovum of the chick (*Gallus domesticus*) for the total embryonic period by growing it in a series of separate culture systems. This is the first report of a complete *in vitro* method for a homoiothermic animal. The technique opens the way to the investigation of developmental events in birds that require access to the embryo at the single-cell stage, and in particular to the genetic manipulation of the fertilized ovum.

Chick development has been divided into three periods for present purposes: fertilization to blastoderm formation (I) lasts for one day⁶, embryogenesis (II) lasts for three days and embryonic growth (III) for eighteen days⁷. Fertilization takes place in the anterior oviduct, after which the yolk-laden ovum is encased in albumen secreted by the magnum. Around the time of the first division of the zygote, some 4.5 h after ovulation, the shell membrane is deposited in the isthmus and the albumen

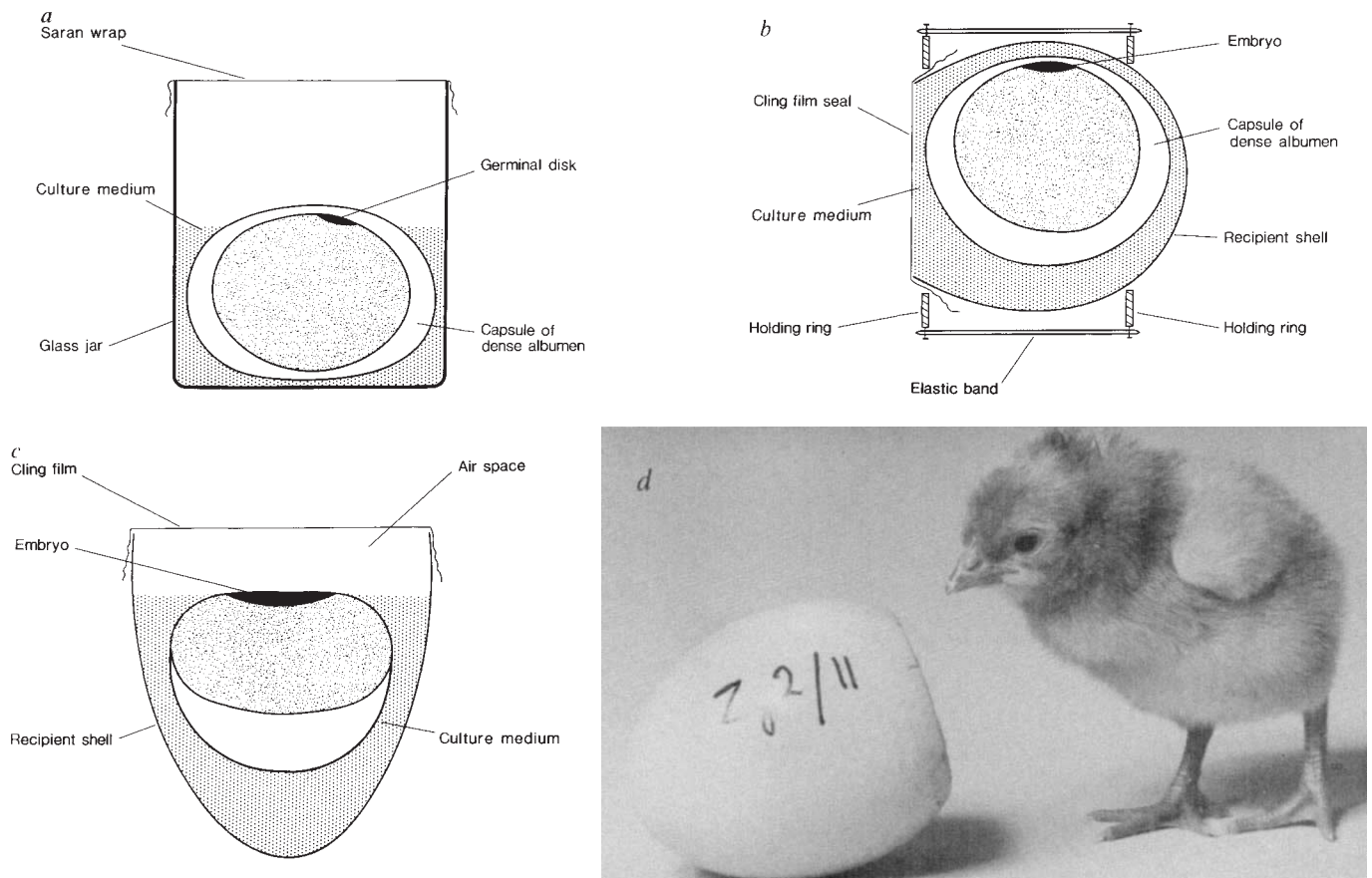


Fig. 1 *a*, Culture system (I) for the oviductal phase of chick development from fertilized ovum to blastoderm formation (day 0–day 1). *b*, Culture system (II) for the phases of embryogenesis and early growth in the chick (day 1–day 6). *c*, Culture system (III) for the growth phase of chick development (day 4–day 22). *d*, A 2 day-old chick reared in culture from fertilized ovum to hatch.

Methods. *a*, Hens (Warrens) on a daily reproductive cycle at 28–32 weeks of age were artificially inseminated (Rhode Island Reds) at weekly intervals. Fertilized ova of broadly similar age were recovered from the oviducts of hens killed at 2.75 h after the preceding egg was laid. The ova were located in the magnum, at a distance of 50–150 mm from its border with the isthmus and were surrounded by a partially formed capsule of dense, viscous albumen. At this stage, the pronuclei are undergoing enlargement in preparation for syngamy in the germinal disk of the ovum¹⁹. The ova were placed in glass jars (60 ml) and medium (8–16 ml) was added to a level in line with the germinal disk, but below the upper surface of the albumen capsule. The volume of medium employed depended on the size of the capsule, which increases as the ovum transverse the magnum. The culture medium consisted of liquid albumen collected from freshly laid eggs (3 parts) and a salt solution (2 parts). The salt solution contained 50 mM K HCO₃, 30 mM NaHCO₃, 10 mM KCl, 2.5 mM Mg Cl₂ · 6H₂O, 0.7 mM CaCl₂ · 2H₂O and 11 mM glucose. The pH of the medium was lowered from an initial value of 8.4 to 7.2–7.4 by stirring in an atmosphere of CO₂. After sealing the jars with saran wrap, secured with an elastic band, the preparations were incubated at 41–42°C, for 24 h. Sterile precautions were taken during all culture procedures. *b*, Recipient shells were prepared from freshly laid eggs by drilling off the sharp end of the shell and rinsing with distilled water. For embryos cultured from day 0, the volume of the recipient shell was equal to that of the preceding egg laid by the donor hen. After transferring the contents of the jar in system I to the recipient shell, the shell was filled with medium (liquid albumen, 2 parts; salt solution, 1 part), pH 8.2–8.4, and sealed with cling film secured by nylon rings. For embryos transferred from unincubated eggs (day 1), the recipient shells were equal in volume to the contents of the donor egg. The shells were filled with undiluted albumen, if required. The reconstituted eggs were rocked intermittently through an angle of 90° in hourly cycles at 38°C, relative humidity (RH) 40–50%, in an automatic incubator (Curfew Model 247). For the total-culture method, the preparations were incubated for 3 days and for the short-term method they were incubated for 6 days. *c*, Recipient shells were prepared from the double-yolked eggs of a commercial broiler strain. The blunt end of the shell was removed, making an opening of 40–45 mm in diameter, and the shell rinsed out with distilled water. For embryos maintained in culture from day 0, the recipient shells were 15–18 ml larger in volume than the preceding eggs laid by the donor hens. For embryos maintained in culture from day 1, eggs selected for host shells were 27–30 g heavier than the donor eggs. After transferring the contents of the reconstituted eggs of system II to the larger shells, a double layer of cling film was placed over the aperture and glued to the shell with albumen containing penicillin and streptomycin. The cultures were incubated at 38°C, RH 40–60%, and rocked intermittently through an angle of 30° in hourly cycles for 5 days, then maintained in a stationary position for the remaining period. The temperature was lowered by 2°C for the final 3 days. Perforations were made in the cling film cover at 1–2 days before hatching, and the cling film replaced with a lid some 12 h before hatch.

is doubled in volume by the absorption of uterine fluid. In the final 18 h of the oviductal phase, the shell is calcified. The second and third phases take place in the shelled egg. Three discrete culture systems have been devised to meet the changing demands at successive stages of development, and the embryos transferred from one system to the next at appropriate times.

The first system covered the oviductal phase and is described in Fig. 1*a*. After incubation for 1 day, the formation of a blas-

toderm was indicated by the appearance, in the germinal region, of an opaque ring (2 mm in diameter) surrounding a semi-transparent area. For an accurate assessment of the conditions for growth in this phase, however, it was necessary to culture the embryos through the second phase.

Transfer of the embryos to recipient shells (Fig. 1*b*) was required for development in the phase of embryogenesis. This culture system (II) was a modification of a method⁸ for the

Table 1 Development of fertilized ova (day 0) and blastodermal embryos (day 1) of the chick in a series of three culture systems

Culture system [duration, days (d)]	Number of cultures		day 4	Live embryos/total cultures (%)			Chicks/total cultures (%) days 22-23
	day 0	day 1		day 7	day 9	day 19	
I (1d) to II	43		ND	67	0	0	0
II		47	ND	83	60	0	0
I (1d) to II (3d) to III	71		76	46	37	21	7
II (3d) to III		49	87	53	41	35	22

The data include healthy and weak chicks. ND, not determined.

* The embryos were staged in days of development from fertilization. Their developmental age, from day 1 onwards, corresponded with the stages given in the table of Hamburger and Hamilton⁷.

short-term culture of quail embryos and it was effective in extending the course of chick development into the early growth phase. After transfer from system I and a further 6 days of incubation, the survival rate was around 60% (Table 1). Control embryos obtained from newly-laid eggs (day 1) and cultured in recipient shells had a survival rate of 80% at 6 days. In each group, there was a sharp increase in mortality during days 8 and 10 of development, respectively.

For further development, embryos cultured in systems I and II were transferred after 4 days' incubation, to large recipient shells containing an air space (Fig. 1c). The method of transfer and the culture system were based on procedures⁵ devised for chick embryos from 3-day incubated eggs. About half of the cultures were lost or slightly damaged by the manipulation and did not survive beyond day 9 (Table 1). Of the remaining embryos, about one fifth hatched on day 22 or 23 of culture. Two chicks were healthy (Fig. 1d) and three were weaklings, having either a leg deformity or partially retracted yolk sacs. Control embryos cultured from day 1 in systems II and III had a higher hatch rate than those cultured from day 0, and about half of the hatchlings were healthy. Some reached maturity at the expected age of 18 weeks, and have produced viable offspring.

The method for system I, covering the oviductal phase, was designed to provide an environment similar to that *in vivo*. Therefore, the culture vessel was made relatively air-tight by sealing it with a material, saran wrap, that has low gaseous permeability⁹. Loss of CO₂ from the chamber was minimized, and, in turn, the pH of the culture medium was maintained at a level compatible with the pH of the albumen¹⁰ in the oviduct. Within a few hours of lay, the pH of the egg albumen rises from 7.4 to 8.4, due to the efflux of CO₂ through the shell¹¹. The composition of the salt solution in the culture medium was based on data^{12,13} for the differences in ionic composition of albumen in uterine eggs, before plumping, and in laid eggs, and on data¹⁴ for the glucose concentration in uterine fluid. This solution was mixed with liquid albumen in proportions that

brought the total plumping fluid content (natural and synthetic) in culture systems II and III to the level obtained in the egg. Only a portion of the requisite amount of medium was used in system I, as the addition of the total amount submerged the preparation and resulted in developmental arrest.

The proper relationship of the embryo with the superstratum was also crucial for development in phases II and III. In contrast to the procedure in system I, in system II the albumen capsule overlying the embryo was subjected to repeated bathing with the medium. It has been shown that the blastoderm develops on intact yolk only when the egg contents are forcibly submerged in liquid albumen¹⁵. Development in the second phase is associated with the transport of fluid from the albumen into the sub-embryonic cavity¹⁶. An excess of medium above the embryo was not detrimental in the third phase until day 8, when the embryos died unless the vascular extra-embryonic membranes were exposed to the atmosphere, as in system III. An essential feature of this system concerns the formation of a normal chorioallantoic membrane and its role in calcium absorption from the shell^{5,17}. The cling-film closure was designed to aid the regulation of the gaseous composition of the air space. This material is partially permeable to O₂, CO₂ and water vapour, and has been employed in shell-less culture systems for the long-term development of 3-day embryos⁹. In the air cell of the egg, the O₂ tension falls and the CO₂ tension rises with time of incubation¹⁸.

We are developing methods for germ-line transformation in poultry, and are using the techniques for culturing fertilized ova for the first third of the embryonic life span to investigate the fate of exogenous DNA inserted into the germinal disc (H. M. Sang & M. P., unpublished). The results for the total culture system indicate that it should now be possible to rear the manipulated ova to hatching, and hence to maturity.

I thank S. G. Tullett and M. Naito for helpful discussions, L. Adams, K. Beattie and C. Landale for technical assistance, and G. Bulfield for support and encouragement. This technique is subject to UK patent application number 8714426.

Received 28 September; accepted 16 November 1987.

- Howarth, B. *Biol. Reprod.* **3**, 338-341 (1971).
- Kochav, S. & Eyal-Giladi, H. *Science* **171**, 1027-1029 (1971).
- Olzanska, B., Szolajski, E. & Lassota, Z. *Wilhelm Roux Arch. dev. Biol.* **193**, 108-110 (1984).
- New, D. A. T. *The Culture of Vertebrate Embryos* (Academic, London, 1966).
- Rowlett, K. & Simkiss, K. *Br. Poult. Sci.* **28**, 91-101 (1987).
- Kochav, S., Ginsburg, M. & Eyal-Giladi, H. *Dev. Biol.* **79**, 296-308 (1980).
- Hamburger, V. & Hamilton, H. L. *J. Morph.* **88**, 49-67 (1951).
- Callebaut, M. *Poult. Sci.* **62**, 1657-1659 (1983).
- Dunn, B. E., Fitzharris, T. P. & Barnett, B. D. *Anat. Rec.* **199**, 33-43 (1981).

- Sauveur, B. & Mongin, P. *Ann. Biol. anim. Biochem. Biophys.* **11**, 213-224 (1971).
- Dawes, C. M. *Biol. Rev.* **50**, 351-371 (1975).
- Beadle, B. W., Conrad, R. M. & Scott, H. M. *Poult. Sci.* **17**, 498-504 (1938).
- Leonard, E. M. thesis, Univ. Edinburgh (1968).
- Davidson, M. F. & Draper, M. H. *J. Physiol., Lond.* **202**, 119-120P (1969).
- Romanoff, A. L. *Anat. Rec.* **87**, 365-369 (1943).
- New, D. A. T. *J. Embryol. exp. Morph.* **4**, 221-227 (1956).
- Ono, T. & Wakasugi, N. *Poult. Sci.* **63**, 159-166 (1984).
- Paganelli, C. V. & Rahn, H. in *Respiration and Metabolism of Embryonic Vertebrates* (ed. Seymour, R. S.) 193-204. (Junk, Dordrecht, 1984).
- Perry, M. M. *J. Anat.* **150**, 99-109 (1987).

Regulation of segment polarity genes in the *Drosophila* blastoderm by *fushi tarazu* and *even skipped*

P. W. Ingham*, N. E. Baker† & A. Martinez-Arias‡§

* Imperial Cancer Research Fund, Developmental Biology Unit, Department of Zoology, South Parks Road, Oxford OX1 3PS, UK.

† Department of Biochemistry, University of California, Berkeley, California, 94720, USA

‡ Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

During the late cellular blastoderm stage of *Drosophila* embryogenesis the segmentation genes *engrailed*, *en*, and *wingless*, *wg*, become expressed in two series of 14 stripes¹⁻⁴ which will subsequently coincide with the anterior⁵ and posterior⁴ limits of each parasegment⁶. Previous studies⁷⁻¹⁰ have shown that the establishment of the pattern of *en* stripes depends upon the activity of the homoeobox-containing pair-rule genes¹¹ *fushi tarazu*, *ftz*¹² and *even skipped*, *eve*^{8,13}. Here we show that these two genes also control the spatial expression of *wg*. Whereas *ftz* and *eve* behave as activators of *en* we find that both genes are required to repress *wg* expression, so that *wg* becomes expressed only in the narrow stripes of cells which come to separate the *ftz* and *eve* bands at the end of the blastoderm stage. In contrast, we propose that the precise positioning of the *en* stripes depends upon signals generated in a combinatorial manner¹⁴ by the overlaps between the *ftz* or *eve* domains and those of other pair rule genes, specifically *odd* paired, *opa*¹⁵ and *paired*, *prd*^{11,16,17}.

In the early *Drosophila* embryo, the transcripts of several pair-rule genes accumulate in a series of rapidly evolving and partially overlapping stripes along the antero-posterior body axis^{8,9,17-19}. Although some of these genes encode functions primarily involved in the refinement of these patterns^{7,20,21} others appear to play a more direct role in establishing the final subdivision of the embryo into parasegments. Good candidates for the latter class are the *ftz* and *eve* genes both of which encode homoeo-domain proteins^{8,12,13} and act as regulators of *en* expression⁷⁻¹⁰. By the end of the blastoderm stage, transcripts of both genes are present in a series of regularly spaced stripes, three nuclei wide, which are precisely out of phase with one another^{8,9}. The evolution of both patterns requires the function of the two 'primary' pair-rule genes *hairy* and *run1*^{7,21}. As the stripes of these and other pair-rule genes undergo a continuous narrowing^{8,9,18,19} it is not clear when and where their expression is critically required. It has recently been suggested that only the anterior margins of the *eve* and *ftz* stripes have an instructive role²², serving to define the boundaries of parasegments by the activation⁷⁻⁹ of the *en* gene. Such a model raises the question as to how the precise domains of other segment polarity genes, such as *wg*⁴, are defined in the blastoderm. To investigate this question we have analysed the effects of *ftz* and *eve* mutations on the activation of *wg* expression.

Expression of *wg* in mutant *ftz* and *eve* blastoderms was monitored by hybridizing tritiated antisense *wg* RNA to sections of embryos derived from parental flies heterozygous for transcript-minus mutant alleles of *ftz* (*ftz*^{w20} and *Df* (3R) 4Scb) or *eve* (*Df* (2R) *eve*^{1,27}). Mutant embryos were identified by hybridizing adjacent sections with labelled *ftz* or *eve* probes, respectively. Typical patterns of *wg* expression in both mutants are shown in Fig. 1. In both cases there is a reduction in the number of *wg* stripes established; most of those remaining are significantly wider than in wild type, in most cases spanning five cells rather than one. The simplest way of explaining how such patterns might arise is to suppose that each novel broad

stripe is generated by filling in between consecutive pairs of normal *wg* stripes. In the case of *ftz*, this would occur between stripes 1 and 2, 3 and 4, 5 and 6 and so on, and in the case of *eve*, between stripes 0 and 1, 2 and 3, 4 and 5 and so on. (See Fig. 1d and e). This 'filling in' would be due to the failure to repress *wg* expression within the normal *ftz* or *eve* domains in the absence of either gene. The exceptionally broad anteriormost stripe seen in *eve* embryos would presumably reflect not only the requirement for *eve* repression between alternate pairs of *wg* stripes but also the premature decay of the first *ftz* band seen in *eve* embryos^{9,20}.

According to this interpretation, *wg* would normally only be expressed in the regions between the *eve* and *ftz* bands, which, at the end of the cellular blastoderm stage, are three cells wide and separated by one cell^{8,9,13}. To test this interpretation we have mapped the *wg* domains with respect to the *eve* domains in *ftz*⁻ embryos. As *eve* expression is independent of *ftz* function⁹ we would expect that within the metamer region of *ftz*⁻ embryos all cells will express either *eve* or *wg*. An example of a section of a *ftz*⁻ embryo hybridized with *wg* and *eve* probes is shown in Fig. 2c. As expected, all the cells in the metamer region are labelled, the signal due to hybridization to *wg* transcript being weaker than that due to hybridization to *eve* transcript.

If each *wg* domain coincides with the posterior limit of each parasegment at the cellular blastoderm stage, as it does in extended germ-band embryos⁴, then it follows from these results that the anterior margin of each *ftz* and *eve* transcriptional domain should be precisely in register with the anterior border of each parasegment. Such an inference is consistent with the finding that, in the extended germ-band embryo, the stripes of cells which express *en* protein coincide with the anteriormost cells expressing the β -galactosidase gene under the control of the *ftz* or *eve* promoters²². To confirm this, we hybridized wild-type blastoderm embryos with a mixture of *ftz* and *en* probes. We find that the *en* stripes, which define the anterior margin of each presumptive parasegment, do indeed coincide with the anterior margins of the *ftz* stripes (Fig. 2d).

Previous studies have shown that the activation of the even-numbered *en* stripes depends on *ftz*⁺ function⁷. By contrast, *eve*⁺ function is required for the establishment of both the odd-numbered and even-numbered *en* stripes^{8,9} (see Fig. 3a). This latter function of *eve* can be ascribed to its expression in seven additional minor stripes which lie within the *ftz* domains and first become visible at the end of the blastoderm stage^{8,13}. There are, however, *eve* alleles which retain some activity and remove only the odd-numbered *en* stripes¹⁰, suggesting that the principal function of the gene in its major domains of expression is analogous to that of *ftz*. Such a congruity is supported by our finding that both genes regulate the spatial expression of *wg* in an analogous fashion.

It has been suggested that the main function of the *ftz* and *eve* stripes is the demarcation of the parasegment boundary and that expression of either gene away from this boundary may be immaterial²². Our finding that removal of *ftz* or *eve* activity leads to the de-repression of *wg* in all nuclei which normally express these genes argues strongly against this view. Also it is clear that the dynamic properties of the *ftz* and *eve* expression patterns are crucial to the function of these genes, the narrowing of each pair-rule stripe from four to three nuclei being the necessary condition for the establishment of the *wg* domains. By contrast, *en* is activated within the *ftz* or *eve* stripes, but in only one of the three nuclei which express either pair-rule gene (see Fig. 3b). This suggests either that *en* responds to a threshold value of *eve* or *ftz* activity within their respective domains or alternatively, that the *en* stripes are specified by the combined expression of *ftz* or *eve* and some other pair-rule gene activities (see ref. 7). A good candidate for generating such a combinatorial signal is the *paired*, *prd*, gene^{11,16,17}, because absence of *prd* function results in the elimination of the odd-numbered

§ Present address: Department of Zoology, Downing Street, Cambridge CB2 3EJ, UK.

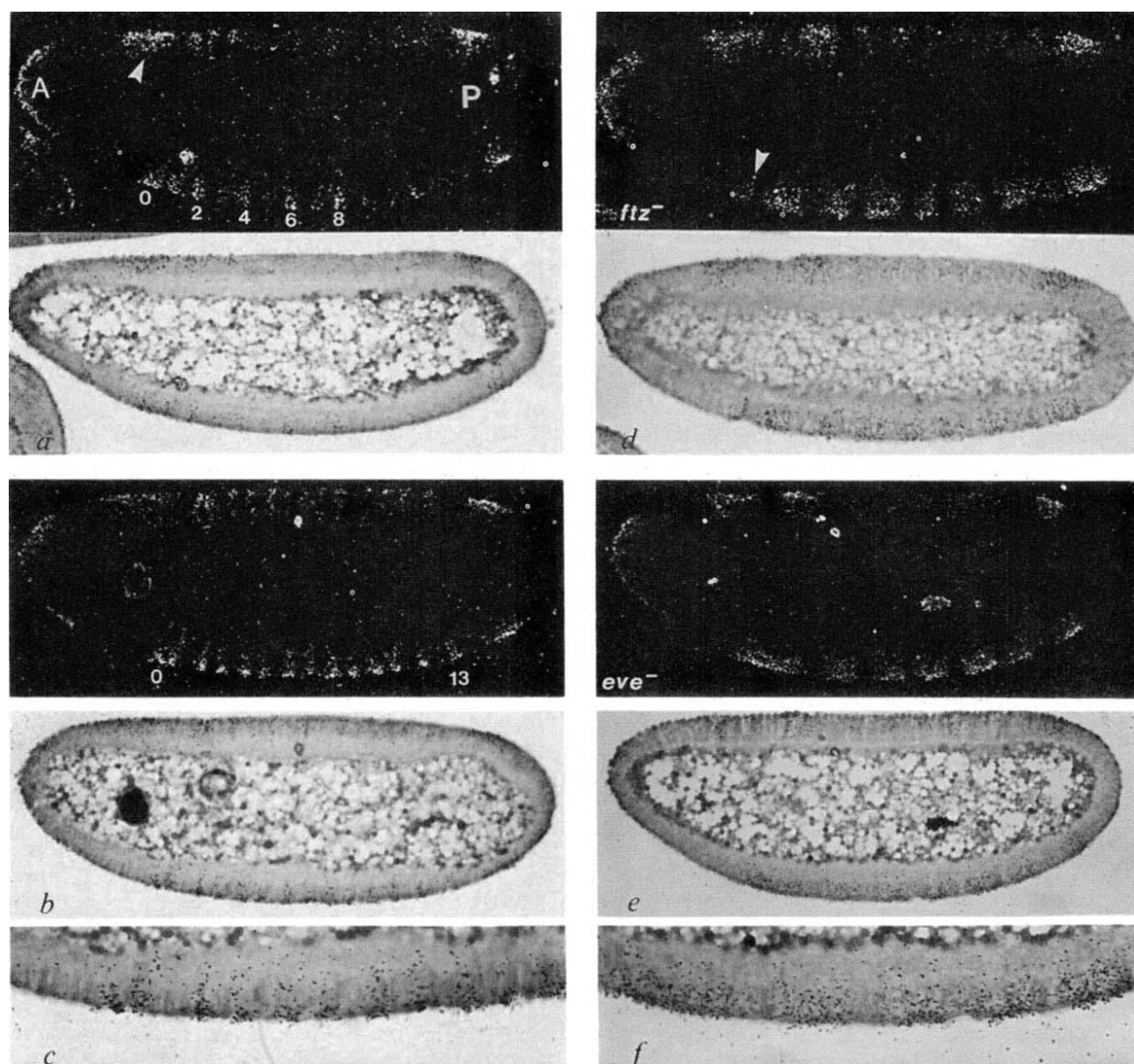


Fig. 1 Dark-field and bright-field images of wild-type (*a-c*) and mutant (*d-f*) embryos. The onset of *wg* expression in the wild-type blastoderm (*a*) is characterized first by the appearance of two sites of transcript accumulation: around the anterior pole (A), and in a broad stripe near the posterior pole (P). Subsequently a third region of accumulation is established antero-dorsally (arrowed). This expression precedes the generation of 14 stripes throughout the presumptive metameric region. The stripes, numbered 0-13 to indicate their parasegmental location, appear in an antero-posterior progression, the odd-numbered ones appearing slightly in advance of their even-numbered neighbours (see *a*). By the onset of gastrulation (*b*) 14 stripes are present between 15% and 75% EL (EL—egg length 0%—posterior pole). Each stripe spans about one nucleus (see detail in Panel *c*). In *ftz*^{W20}/*Df*(3R) 4*Sch* embryos (*d*) only the most anterior stripe (arrowed) is of normal width. Posterior to the cephalic fold, seven broad stripes of *wg* expression form. Each of these spans about five nuclei. *e*, Expression of *wg* in a *Df*(2R) *eve*^{1.27} blastoderm. As in the case of the *ftz* mutants, fewer but broader stripes of *wg* expression are present. In this case the most anterior stripe is very broad, spanning 12-13 nuclei. This region corresponds to that demarcated by stripes 0 and 3 in the wild-type. Posterior to this are a further five stripes, each around five nuclei wide (see detail in Panel *f*). An identical pattern of expression is seen in embryos hemizygous for the allele *eve*^{R13} (data not shown).

Methods. Embryos were fixed, embedded and sectioned as previously described¹⁹. Single-stranded RNA probe, labelled with ³H, was prepared from the T3 promoter of plasmid pwg-c14a (ref. 4), and hybridized to the sections as described¹⁹. Mutant embryos, were identified by hybridizing adjacent sections with probes for either the *ftz* or *eve* transcript which have been described elsewhere^{7,8}.

en stripes^{10,23} (see Fig. 3*a*). The expression pattern of *prd* in the late cellular blastoderm differs from that of *ftz* or *eve*, being expressed in 14 bands each two cells wide (with the exception of the most posterior band)¹⁷. Double labelling experiments have suggested that alternate *prd* stripes lie adjacent and posterior to each *ftz* stripe¹⁷. As each *prd* stripe is two cells wide, it follows that alternate stripes will overlap the anterior part of each *eve* band by one cell (see Fig. 3*b*). Thus the combination of *eve* and *prd* expression could serve to define the position of the odd-numbered *en* stripes that require the activity of both genes for their establishment. By symmetry, we suggest that a similar signal is provided by the combined expression of *ftz* and

the pair-rule gene *opa*, to establish the even-numbered *en* stripes. Indeed, absence of *opa*⁺ function results in the elimination of these stripes^{23,24} (see Fig. 3*a*) and the *opa* phenotype is approximately the reciprocal of that of *prd*²⁵. We therefore surmise that *opa* will exhibit a reciprocal pattern of expression (see Fig. 3*b*). These patterns of expression give rise to at least two other combinations of gene activities, namely *ftz* + *prd* and *eve* + *opa*, which may be used to specify the domains of expression of other segment polarity genes. Absence of *prd*⁺ or *opa*⁺ function also results in the elimination of alternate stripes of *wg* expression (data not shown; see Fig. 3*a*). This function of *prd* and *opa* appears redundant with respect to the specification of

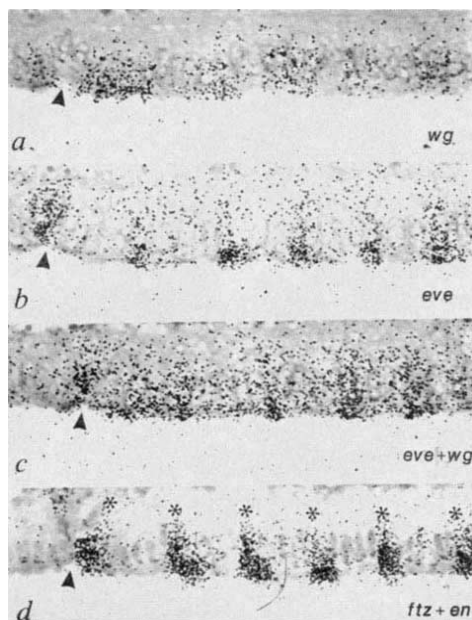


Fig. 2 *a*, Part of a saggital section of a *ftz*^{w20}/*Df*(3*R*)4*Scb* embryo hybridized with *wg* probe showing the broad bands of *wg* transcript accumulation. The adjacent section of this embryo was hybridized with *eve* probe (*b*), revealing the seven-stripped pattern of *eve* expression, which is independent of *ftz*⁺ function. The next section (*c*) was hybridized with a mixture of *wg* and *eve* probes. All of the cells are labelled, indicating that the broad stripes of *wg* expression lie in between the *eve* stripes and therefore correspond to the regions where *ftz* would normally be expressed. Arrowheads, position of the cephalic fold. *d*, The relationship between the *ftz* and *en* transcriptional domains. A saggital section through the ventral region of an embryo which has just begun gastrulation. Arrow head, cephalic fold. The section was hybridized with a mixture of tritiated *en* and *ftz* probes. The grains due to hybridization of the *en* probe extend deep into the cytoplasm of the cells whereas the *ftz* signal is restricted to the periphery. At this stage only the even-numbered *en* bands can be clearly visualized. The anterior margin (to the left) of each *ftz* domain can be seen to coincide with the even-numbered *en* stripes (*).

the position of *wg* expressing cells in the blastoderm. Such a requirement may, however, be crucial to delimiting the domains of *wg* expression. If *wg* were exclusively under negative control, the *wg* domain might become progressively broader with time because the *eve* and *ftz* domains both continue to narrow as gastrulation proceeds. The requirement for *opa* and *prd* would thus serve to ensure that the release of *wg* from repression be restricted to the first cells to stop expressing *ftz* and *eve* (see Fig. 3*b*). Alternatively it is possible that the maintenance of the *wg* domains after the blastoderm stage depends upon regulatory interactions between other segment-polarity genes.

In terms of the foregoing formal genetic analysis, *ftz* and *eve*, both of which encode homoeo-domain proteins, act as positive regulators of *en* but negative regulators of *wg*. Also, *ftz* is required for the initial modulation of the homoeotic genes *Scr*,

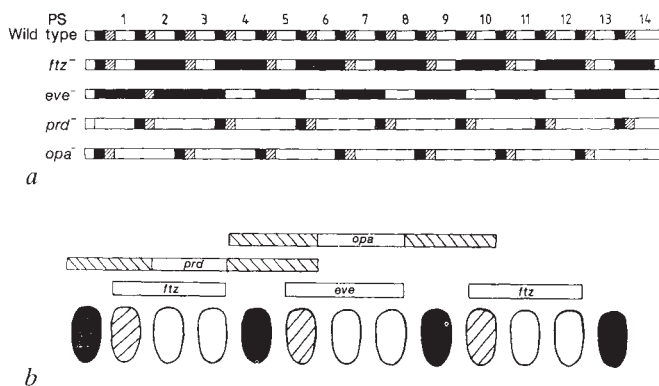


Fig. 3 *a*, Schematic representation of the patterns of *en* and *wg* expression in the wild-type cellular blastoderm and in four pair-rule mutants. Hatched bars represent *en* domains, solid bars *wg* domains. Absence of *eve* or *ftz* expression results in the expression of *wg* throughout the normal *eve* or *ftz* domains. In the case of *eve* there is an additional de-repression of *wg* between the first and second *eve* domains, presumably due to the repression of the first *ftz* stripe by the *eve* mutation²⁰. In contrast, absence of *prd* or *opa* function results in the elimination of alternate *wg* bands and of the adjacent *en* bands, but the remaining bands are of normal size and position (data not shown). *b*, Schematic representation of part of the periphery of a blastoderm embryo showing the relation of the known *ftz*, *eve* and *prd* transcriptional domains, and a postulated domain of *opa* expression. At the end of the blastoderm stage the *ftz* and *eve* bands have narrowed to an average width of three cells and are separated by a single cell^{8,9}. At the same time each *prd* domain becomes split into two bands each two cells wide, by the elimination of transcript from the middle (open bar) region of each domain¹⁷. The repression of *wg* by *ftz* and *eve* would allow *wg* expression only in the cells indicated (solid nuclei). The overlap of the *prd* and *eve* domains identifies the cell which will express *en* (hatched nuclei). That this combination specifies the odd-numbered *en* stripes is supported by the finding that their establishment depends both upon *eve* and *prd* activity. By analogy, we suggest that *opa* might be expressed in a similar, though complementary, pattern to *prd*. Thus the posterior of each *opa* domain would overlap the anterior of each *ftz* domain, thereby specifying the activation of the even-numbered *en* stripes. According to this scheme *prd* and *opa* will ultimately be expressed in identical stripes corresponding to each *en* and *wg* domain. Thus the requirement for *opa* and *prd* to activate alternate *wg* stripes is dependent upon context. The nature of this context is obscure.

Antp and *Ubx*²⁴. The specificity of each of these functions of the *eve* and *ftz* products will depend on their interaction with other gene products at particular given promoters. For example, in the case of homoeotic genes, gap-gene products might play an important role²⁶ whereas in the case of segment polarity genes, we suggest the specificity is set by other pair-rule genes. A situation in which the same DNA-binding protein behaves as an activator or repressor of different promoters has recently been described in detail for RAP-1 protein in yeast²⁷.

We thank P. A. Lawrence for discussions and M. E. Akam for many helpful suggestions. This work was supported by the ICRF and the MRC of Great Britain.

Received 30 September; accepted 2 December 1987.

1. Fjose, A., McGinnis, W. J. & Gehring, W. G. *Nature* **313**, 284-289 (1985).
2. Kornberg, T., Siden, I., O'Farrell, P. & Simon, M. *Cell* **40**, 45-63 (1985).
3. DiNardo, S., Kuer, J. M., Theis, J. & O'Farrell, P. H. *Cell* **43**, 59-69 (1985).
4. Baker, N. E. *EMBO J.* **6**, 1765-1773 (1987).
5. Ingham, P. W., Martinez-Arias, A., Lawrence, P. A. & Howard, K. R. *Nature* **317**, 634-636 (1985).
6. Martinez-Arias, A. & Lawrence, P. A. *Nature* **313**, 639-642 (1985).
7. Howard, K. R. & Ingham, P. *Cell* **44**, 949-957 (1985).
8. Macdonald, P., Ingham, P. & Struhl, G. *Cell* **47**, 721-734 (1986).
9. Harding, K., Rushlow, C., Doyle, H. J., Hoey, T. & Levine, M. *Science* **233**, 953-959 (1986).
10. Martinez-Arias, A. & White, R. A. H. *Development* (in the press).
11. Nusslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795-801 (1980).
12. Laughon, A. & Scott, M. P. *Nature* **310**, 25-31 (1984).
13. Frasch, M., Hoey, T., Rushlow, C., Doyle, H. & Levine, M. *EMBO J.* **6**, 749-759 (1987).

14. Gergen, J. P., Coulter, D. & Weischaus, E. F. in *Gametogenesis and the early Embryo* (ed. Gall, J.) 195-220 (Liss, New York, 1986).
15. Jurgens, G., Wieschaus, E., Nusslein-Volhard, C. & Kluding, H. *Roux Arch. Dev. Biol.* **193**, 283-295 (1984).
16. Sander, K., Lohs-Schardin, H. & Burmann, H. *Nature* **287**, 841-843 (1980).
17. Kilcher, F., Baumgartner, S., Bopp, D., Frei, E. & Noll, M. *Nature* **321**, 493-499 (1986).
18. Hafen, E., Kuroiwa, A. & Gehring, W. J. *Cell* **37**, 833-841 (1984).
19. Ingham, P. W., Howard, K. R. & Ish-Horowitz, D. *Nature* **318**, 439-445 (1985).
20. Carroll, S. B. & Scott, M. P. *Cell* **45**, 113-119 (1986).
21. Ingham, P. W. & Gergen, J. P. (in preparation).
22. Lawrence, P. A., Johnston, P., Macdonald, P. & Struhl, G. *Nature* **328**, 440-442 (1987).
23. DiNardo, S. & O'Farrell, P. *Genes Dev.* (in the press).
24. Ingham, P. W. & Martinez-Arias, A. *Nature* **324**, 592-597 (1986).
25. Nusslein-Volhard, C., Kluding, H. & Jurgens, G. *Cold Spring Harb. Symp. Vol. L*, 145-154 (1985).
26. Ingham, P. W., Ish-Horowitz, D. & Howard, K. R. *EMBO J.* **5**, 1659-1665 (1986).
27. Shore, D. & Nasmyth, K. *Cell* (in the press).

HIV infection is blocked *in vitro* by recombinant soluble CD4

Richard A. Fisher*, Jeanne M. Bertonis*, Werner Meier*, Victoria A. Johnson†, Donna S. Costopoulos*, Theresa Liu*, Richard Tizard*, Bruce D. Walker†, Martin S. Hirsch†, Robert T. Schooley† & Richard A. Flavell*

* Biogen Research Corporation, 14 Cambridge Center, Cambridge, Massachusetts 02142, USA

† Infectious Disease Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

The T-cell surface glycoprotein, CD4 (T4), acts as the cellular receptor for human immunodeficiency virus, type 1 (HIV-1), the first member of the family of viruses that cause acquired immunodeficiency syndrome¹⁻⁴. HIV recognition of CD4 is probably mediated through the virus envelope glycoprotein (gp120) as shown by co-immunoprecipitation of CD4 and gp120 (ref. 5) and by experiments using recombinant gp120 as a binding probe⁶. Here we demonstrate that recombinant soluble CD4 (rsT4) purified from the conditioned medium of a stably transfected Chinese hamster ovary cell line is a potent inhibitor of both virus replication and virus-induced cell fusion (syncytium formation). These results suggest that rsT4 is sufficient to bind HIV, and that it represents a potential anti-viral therapy for HIV infection.

Radiolabelled synthetic oligonucleotide hybridization probes were used to isolate a full length complementary DNA copy of CD4 (Fig. 1) from a λ gt10 library synthesized from poly(A)⁺ messenger RNA isolated from human peripheral blood lymphocytes (PBLs) (ref. 7). DNA sequence analysis of the full-length (3.06 kilobases) PBL cDNA reveals that the CD4 open-reading frame encodes three amino-acid substitutions compared to the sequence deduced by Maddon and coworkers⁸. This heterogeneity, found at positions 3 (Asn to Lys), 64 (Trp to Arg) and 231 (Phe to Ser), appears to represent allelic polymorphism. We have modified the three codons at positions 3, 64 and 231 using *in vitro* mutagenesis, resulting in expression of a CD4 protein sequence identical to that previously reported.

Using site-directed mutagenesis to insert in-frame translational stop codons, we constructed three truncated CD4 genes that lack the transmembrane and cytoplasmic domains (Fig. 1).

Fig. 1 Construction of recombinant soluble CD4 mutants.

a, Hydropathy analysis of our full-length soluble CD4 using a computer program Pepplot (University of Wisconsin Genetics Computer Group)¹⁸. *b*, Protein domain structure of full-length CD4 (ref. 8), in which S represents the secretory signal sequence, V the immunoglobulin-like joining region sequence, U the unique, extracellular region sequence, TM the transmembrane sequence and C the cytoplasmic region sequence. The transmembrane amino-acid sequence and some flanking sequence is shown under the TM domain. The underlined amino acids define N- and C-terminal ends of the putative TM domain. *c*, Protein domain structure of recombinant soluble CD4 mutants rsT4.1 in BG377, rsT4.2 in BG380 and rsT4.3 in BG381.

Methods. A peripheral blood lymphocyte (PBL) cDNA library in λ gt10 (ref. 7) was screened in parallel by hybridization with three ³²P-labelled synthetic oligonucleotide anti-sense probes, probe 3 (3'-ACGGAGTGGCCAAAGTCTTCTGTACA-5'), probe 6 (3'-CACGTTAACGAT-CACAAGCCTAACTGA-5'), and probe 9 (3'-TCCGTGAACGAAGACCACGACG-5'). About 10⁶ plaques were screened as described¹⁹, and a full length CD4 clone containing an insert of 3,064 base pairs was isolated. The sequence of the insert was determined by the chemical modification technique²⁰ after it was excised with *Eco*RI and inserted into plasmid pBG312 (ref. 9), to create p-170-2. The CD4 coding sequence of p-170-2 contains three amino-acid-coding substitutions [position 3 (Asn to Lys), position 64 (Trp to Arg), and position 231 (Phe to Ser)]. These three codons were modified in sequential steps using oligonucleotide-directed mutagenesis (positions 3 and 64), or restriction fragment substitution with a fragment including the ser-231 codon of a partial CD4 cDNA isolated from a CD4⁺-lymphocyte cell line²¹ library in λ gt11 (a gift from E. Reinherz). Recombinant soluble CD4 mutants were created by introducing synthetic linkers into the unique *Ava*I site that is 5' to the U/TM domain boundary to produce an in-frame translational stop codon. rsT4.1 contains a stop codon immediately following amino acid residue 362. Similarly, rsT4.2 and rsT4.3 contain stop codons following residues 374 and 377 respectively.

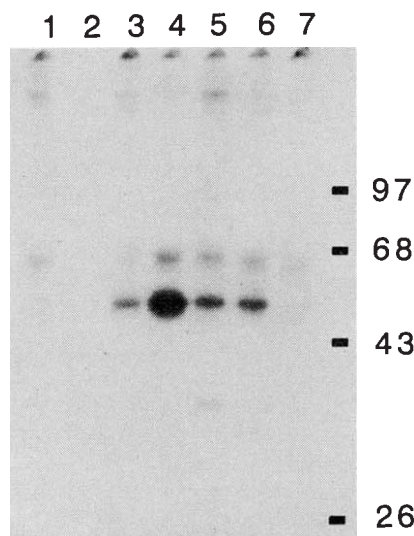
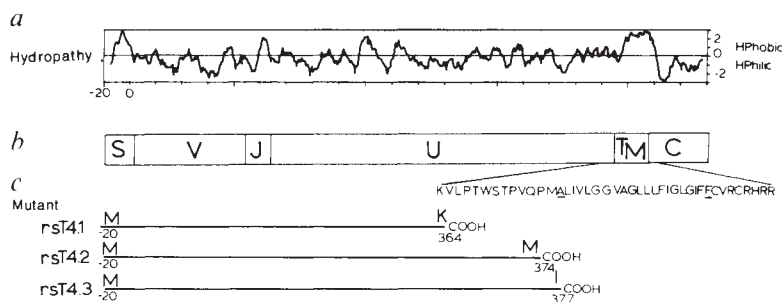


Fig. 2 Detection of rsT4 from the conditioned medium of stably transfected CHO lines. Detection of ³⁵S-metabolically labelled rsT4.2 and rsT4.3 by immunoprecipitation with OKT4. Lane 1, immunoprecipitation from conditioned medium of a CHO clone stably cotransfected with vector pBG312 and pAdD26; lane 2, blank; lanes 3 and 4, immunoprecipitation from conditioned medium of CHO clones stably cotransfected with pBG380 (rsT4.2) and pAdD26; lanes 5 and 6, immunoprecipitation from conditioned medium of cells cotransfected with pBG381 (rsT4.3) and pAdD26; lane 7, immunoprecipitation from conditioned medium of cells cotransfected with pBG379 (rfT4) and pAdD26. *M_r* for standard markers are shown to the right (K).

Methods. About 10⁷ transfected CHO cells were metabolically labelled for 5 h at 37 °C with 180 μ Ci ml⁻¹ [³⁵S]cysteine (DuPont, NEN) in 4 ml RPMI cys⁻ medium (Gibco). After labelling, 1 ml filtered conditioned medium was made 0.5 mM with phenylmethylsulphonyl fluoride, and immunoprecipitated with OKT4 and protein A Sepharose²². Immunoprecipitates were subjected to SDS-PAGE (ref. 23) on 10% polyacrylamide gels. Autoradiography was performed with X-Omat X-ray film (Eastman Kodak).

Each of the three soluble gene constructions (rsT4.1, rsT4.2 and rsT4.3) was inserted into a derivative of the animal cell expression plasmid vector pBG312 (ref. 9) under the transcriptional control of the adenovirus major late promoter, giving the plasmids pBG377, pBG380 and pBG381 respectively. We also

Fig. 3 Purification and characterization of rsT4.2. *a*, SDS-PAGE analysis under reducing conditions of fractions eluted from an anti-CD4 mAb immunoaffinity column. The gel was stained for protein with Coomassie blue. *b*, SDS-PAGE analysis under reducing conditions of pooled peak fractions of immunoaffinity-purified rsT4. M_r (K) of standards are shown to the left. A portion of the pooled peak fractions was subjected to N-terminal sequence analysis. *c*, N-terminal sequence analysis of rsT4.2 and comparison to the predicted N-terminal sequence⁸.

Methods. The stably transfected CHO clone BG380.G was grown to confluency in 21 Cell Factory (Nunc) systems at 37 °C, in α -modified Eagle's medium (MEM) (Gibco) supplemented with 10% fetal calf serum (FCS), 1 mM glutamine, and the antibiotics penicillin and streptomycin (100 μ g ml⁻¹ each). The confluent monolayers were washed free of FCS with α -MEM, and cultured for four days at 37 °C in α -MEM without FCS. Conditioned medium was collected and concentrated on an S-Sepharose fast-flow column (Pharmacia). Proteins eluted with 20 mM Tris-HCl (pH 7.7) and 0.3 M NaCl were pooled, and loaded on to an immunoaffinity column of the 19 Thy anti-CD4 mAb coupled to Affigel-10 (a gift from E. Reinherz). After extensive washing, the bound proteins were eluted with 50 mM glycine-HCl (pH 2.5), 150 mM NaCl, 0.1 ml EGTA, and 5 μ g ml⁻¹ bovine pancreatic trypsin inhibitor, Aprotinin (Sigma), as 0.5 ml fractions. The eluted rsT4.2 six peak fractions were identified by reducing SDS-PAGE. The yield of purified rsT4.2 is ~ 80 μ g l⁻¹ BG380.G conditioned medium. The immunoaffinity-purified rsT4 (~ 15 μ g) was subjected to N-terminal sequence analysis by automated Edman degradation in an Applied Biosystems 470A gas phase sequencer²⁵. The rsT4.2 N-terminal sequence deduced from sequence analysis is compared to the N-terminal sequence predicted from the CD4 cDNA sequence of Maddon *et al.*⁸.

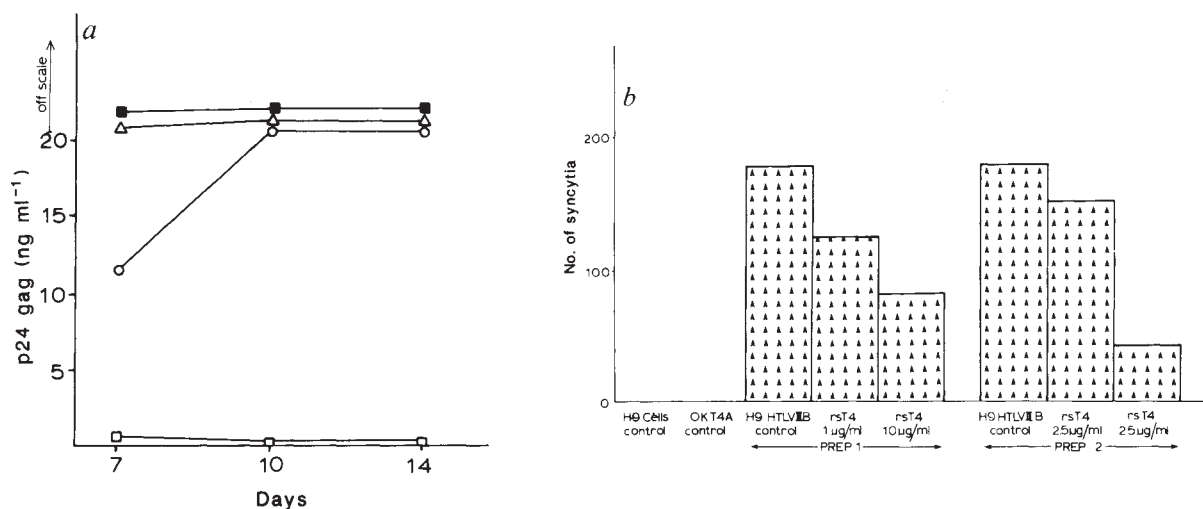
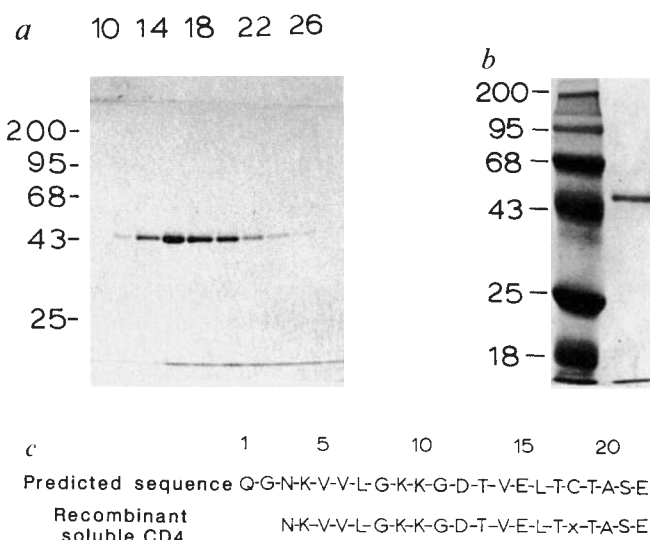


Fig. 4 Antiviral activity of rsT4 measured in HIV-1 replication and cell fusion assays. *a*, Dose-dependent inhibition by rsT4.2 of HIV-1 p24 gag production in infected H9 cells. Concentrations of rsT4.2: Δ , 0.1 μ g ml⁻¹; $\circ$, 1.0 μ g ml⁻¹; $\square$, 10 μ g ml⁻¹; $\blacksquare$, buffer. *b*, Dose-dependent inhibition by rsT4.2 of HIV-induced cell fusion.

Methods. The virus replication assay was carried out essentially as described¹⁴, except that the cultures were propagated in microtitre wells containing 200 μ l. Graded concentrations of rsT4.2 or buffer alone, each in 25 μ l, were preincubated for 1 h at 37 °C in 5% CO₂ with 50 TCID₅₀ (ref. 26) HTLV-III B (a gift from Drs M. Popovic and R. Gallo) in 25 μ l. Following preincubation, H9 cells (1×10^5 cells in 150 μ l RPMI-1640 supplemented with 20% heat-inactivated FCS) were added to the wells, yielding final rsT4 concentrations ranging from 0.1 μ g ml⁻¹ to 10 μ g ml⁻¹. Microtitre plates were incubated at 37 °C in 5% CO₂ for 14 d. Cells were fed by exchanging 100 μ l cell-free supernatant fluid on days 3, 7 and 10 with fresh medium, and no further rsT4 was added during this period. Cell-free supernatant fluid (100 μ l) was analysed for p24 antigen by RIA (Dupont, NEK-040). The C8166 fusion assay has been described²⁷. Two preparations of rsT4 (prep 1 and prep 2) were tested in a 2 h assay. H9 cells (1×10^4), infected chronically with HTLV-III B, were preincubated with varying concentrations of rsT4 in 150 μ l medium in 96-well plates. All assays were done in triplicate. After 1 h incubation at 37 °C in 5% CO₂, 3×10^4 C8166 cells (HTLV-I transformed umbilical cord lymphocytes¹⁵) in 50 μ l were added to the wells. Final well concentrations of rsT4 were 1 μ g ml⁻¹ and 10 μ g ml⁻¹ for prep 1; and 2.5 μ g ml⁻¹ and 25 μ g ml⁻¹ for prep 2. Preincubation with OKT4A (Ortho Diagnostics) at 25 μ g ml⁻¹ served as a control. After plates were incubated for 2 h at 37 °C in 5% CO₂, syncytia (ballooning cytoplasm greater than three lymphocyte cell diameters) were counted. To prevent bias during counting, samples were coded.

made a pBG312-based construction, called pBG379, that directs the expression of recombinant full-length CD4 (rfT4). To test for expression of rsT4 and rfT4, dihydrofolate reductase (DHFR)-mutant CHO cells were transfected with each of the four plasmids and with plasmid pAdd26 (ref. 10) containing the mouse DHFR gene, as described previously¹¹. Stably transfected cell lines were isolated by cloning in selective growth

medium, then screened for rsT4 expression with a CD4 antigen radioimmunoassay (RIA) (ref. 12 and T. Liu, unpublished) and by immunoprecipitation from conditioned media after [³⁵S]cysteine (³⁵S-Cys) metabolic labelling. As shown in Fig. 2, both pBG380 (rsT4.2)- and pBG381 (rsT4.3)-transfected CHO cells make secreted, ³⁵S-labelled proteins that are recognized by OKT4 (ref. 22), and which migrate as molecules of the relative

molecular mass (M_r) predicted for such truncated proteins (49,000; 49K). In contrast, no soluble CD4 antigen could be detected in the conditioned medium of cell lines stably transfected with pBG377 (rsT4.1) or pBG379 (rflT4). Immunoprecipitation analysis of cellular extracts (not shown) suggests that the rsT4.1 gene product is misfolded, which could account for a block in its secretion¹³.

For larger scale production of rsT4, a recombinant CHO clone that secretes rsT4.2, namely BG380.G, was seeded into two cell factory systems (Fig. 3). Medium was collected, concentrated, immunoaffinity purified, and subjected to reducing SDS-PAGE and to N-terminal sequence analysis (Fig. 3) which revealed that the protein purified from BG380.G has an N-terminal sequence corresponding to the predicted CD4 sequence beginning at position 3, asparagine.

Purified rsT4 was tested as an inhibitor of HIV-1 replication in an *in vitro* assay¹⁴. As measured by HIV-1 p24 antigen production, rsT4 effectively blocked virus replication at an initial concentration of $10 \mu\text{g ml}^{-1}$ (Fig. 4a). These p24 antigen RIA data are in agreement with the results obtained from the corresponding cytopathic effect (CPE) assays (data not shown). At the concentrations of rsT4 used, no effects on uninfected H9 cell viability or proliferation were observed.

It has been shown that HIV-1 envelope glycoprotein induces cell fusion of CD4⁺-lymphocytes *in vitro*, resulting in the formation of multinucleated giant cells, or syncytia^{15,16}. The rsT4 preparation was tested for its ability to block CD4-dependent HIV-induced cell fusion. As shown in Fig. 4b, at rsT4 concentrations of either $10 \mu\text{g ml}^{-1}$ or $25 \mu\text{g ml}^{-1}$ (from two separate preparations of rsT4), syncytium formation was inhibited by 54% and 77% respectively.

These experiments could lead to an HIV anti-viral therapy based on the soluble receptor for the virus. As all HIV types appear to use the same or related receptors, rsT4 might achieve broader anti-HIV activity than therapies based on antibodies, which could be strain-specific. This protein may be valuable as a drug in combination with other anti-retroviral agents that block reverse transcriptase, such as azidothymidine (AZT). Primate models using related simian viruses could help test its safety and potential in humans.

We acknowledge the Biogen protein sequencing group for providing rsT4 microsequence analyses, Gary Jaworski for tissue culture assistance, Ellis Reinherz and Peter Sayre for 19Thy 5D7 mAb and advice, and Neenyah Ostrom for preparing the manuscript. In addition, we thank Vicki Sato, Joseph Rosa, Blake Pepinsky, Roy Byington and Teresa Rota for discussion. This work was supported by a Small Business Innovation Research Grant to Biogen, a National Cooperative Drug Discovery Grant for AIDS to M.G.H. and Biogen and by Biogen N.V.

Received 2 October; accepted 19 November 1987.

1. Dalglish, A. G. *et al. Nature* **312**, 763-767 (1984).
2. Klatzmann, D. *et al. Nature* **312**, 767-768 (1984).
3. Maddon, P. J. *et al. Cell* **47**, 333-348 (1986).
4. McDougal, J. S. *et al. J. Immun.* **135**, 3151-3162 (1985).
5. McDougal, J. S. *et al. Science* **231**, 382-385 (1986).
6. Lasky, L. A. *et al. Cell* **50**, 975-985 (1987).
7. Wallner, B. *et al. J. exp. Med.* **166**, 923-932 (1987).
8. Maddon, P. J. *et al. Cell* **42**, 93-104 (1985).
9. Cate, R. L. *et al. Cell* **45**, 685-698 (1986).
10. Kaufman, R. J. & Sharp, P. A. *J. molec. Biol.* **159**, 601-621 (1982).
11. Haynes, J. & Weissmann, C. *Nucleic Acids Res.* **11**, 687-706 (1983).
12. Rao, P. E., Talle, M. A., Kung, P. C. & Goldstein, G. *Cell Immun.* **80**, 310-319 (1983).
13. Gething, M. J., McCammon, K. & Sambrook, J. *Cell* **46**, 939-950 (1986).
14. Ho, D. *et al. J. Virol.* **61**, 2024-2028 (1987).
15. Sodroski, J., Goh, W. C., Rosen, C., Campbell, K. & Haseltine, W. A. *Nature* **322**, 470-474 (1986).
16. Lifson, J. D. *et al. Nature* **323**, 725-728 (1986).
17. Swain, S. L. *Immunol. Rev.* **74**, 129-142 (1983).
18. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
19. Benton, W. D. & Davis, R. W. *Science* **196**, 180-182 (1977).
20. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-580 (1980).
21. Acuto, O. *et al. Cell* **34**, 717-726 (1983).
22. Sayre, P. H. & Reinherz, E. L. *Eur. J. Immun.* **15**, 291-295 (1985).
23. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
24. Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4354 (1979).
25. Pepinsky, R. B. *et al. J. Biol. Chem.* **261**, 4239-4246 (1986).
26. Hartshorn, K. L. *Antimicrob. Agents Chemother.* **31**, 168-172 (1987).
27. Walker, B. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 8120-8124 (1987).

A soluble CD4 protein selectively inhibits HIV replication and syncytium formation

Rebecca E. Hussey*, Neil E. Richardson*, Mark Kowalski†, Nicholas R. Brown*, Hsiu-Ching Chang*, Robert F. Siliciano*, Tatyana Dorfman†, Bruce Walker†, Joseph Sodroski† & Ellis L. Reinherz*

Laboratories of *Immunobiology and †Biochemical Pharmacology, Dana-Farber Cancer Institute and Departments of Medicine and Pathology, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA

The CD4 (T4) molecule is expressed on a subset of T lymphocytes involved in class II MHC recognition^{1,2}, and is probably the physiological receptor for one or more monomorphic regions of class II MHC (refs 1-3). CD4 also functions as a receptor for the human immunodeficiency virus (HIV) exterior envelope glycoprotein (gp120) (refs 4-9), being essential for virus entry into the host cell and for membrane fusion, which contributes to cell-to-cell transmission of the virus and to its cytopathic effects^{10,11}. We have used a baculovirus expression system to generate mg quantities of a hydrophilic extracellular segment of CD4. Concentrations of soluble CD4 in the nanomolar range, like certain anti-CD4 monoclonal antibodies, inhibit syncytium formation and HIV infection by binding gp120-expressing cells. Perhaps more importantly, class II specific T-cell interactions are uninhibited by soluble CD4 protein, whereas they are virtually abrogated by equivalent amounts of anti-T4 antibody. This may reflect substantial differences in CD4 affinity for gp120 and class II MHC.

To produce a secreted form of CD4, the pSP65-T4 complementary DNA was truncated as described in Fig. 1 legend and ligated to a synthetic linker such that either 371 residues (T_{4ex1}) or 370 residues (T_{4ex2}) of the predicted 374 amino acids¹² of the mature extracellular segment would be preserved. After cloning into the pAc373 vector, the truncated CD4 cDNA constructs were integrated into the *Autographa californica* nuclear polyhedrosis virus (AcNPV) genome by homologous recombination¹³. Pure recombinant baculovirus stocks were used to infect *Spodoptera frugiperda* (SF9) cells.

Figure 2a shows that the T_{4ex1} polypeptide is the major secreted product of SF9 cells infected with the T_{4ex1} recombinant baculovirus. The predominant ³⁵S-labelled protein band (45% total labelled material) in supernatants from infected SF9 cells has a relative molecular mass (M_r) of 50,000 (50K) under reducing conditions (Fig. 2a, lane 1) and co-migrates with material immunoprecipitated by an anti-CD4 monoclonal antibody (19Thy5D7) from T_{4ex1} baculovirus-infected SF9 supernatants (Fig. 2a, lane 2) or cell lysates (Fig. 2a, lane 3). The latter also shows a strongly labelled band of 52K, presumably the T_{4ex1} polypeptide plus signal peptide. Although a 50K band is seen in the total cell lysate of T_{4ex1} virus infected cells (Fig. 2a, lane 4), even without immunoprecipitation with anti-CD4 monoclonal antibody, it is a minor component of a mixture of labelled intracellular polypeptides. As expected, no CD4 material was precipitated from supernatants of wild-type AcNPV-infected cells (Fig. 2a, lane 5) or detectable in the total supernatant (Fig. 2a, lane 6).

Large-scale production of soluble CD4 is described in Fig. 2 legend. For each infection, 1 μg purified material was analysed by Coomassie blue staining after SDS-PAGE. As shown in Fig. 2b (lanes 1 and 2), two representative T_{4ex2} preparations yield a protein with M_r 51K (reduced), which we find to be glycosylated using endoglycosidase F experiments (data not shown). The T_{4ex1} protein has M_r 50K (lane 3). These different mobilities are not unexpected as T_{4ex2} contains 17 extra carboxy-terminal amino acids derived from fusion with the polyhedrin

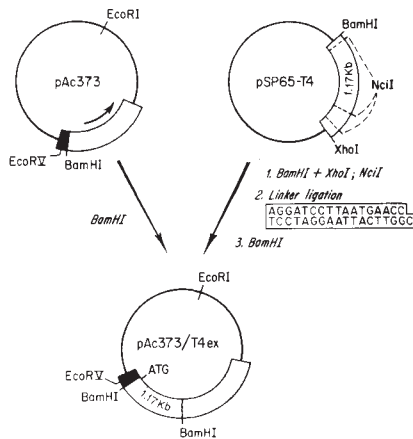


Fig. 1 Construction of baculovirus expression vectors for production of secreted forms of human CD4. The plasmid transfer vector pAc373 contains a single *Bam*HI cloning site 8 bp upstream of the polyhedrin ATG start site. To produce a secreted form of the T4 molecule, the plasmid pSP65-T4 (gift from Dan Littman) was digested with *Bam*HI and *Xho*I to release the CD4 cDNA insert¹² which was subsequently digested with *Nci*I. *Nci*I cleaves CD4 cDNA at nucleotide positions 83, 1,253 and 1,604 yielding a fragment of 1.17 kb which lacks the ATG start codon and terminates just before the transmembrane region. Oligonucleotides 5'AGGATCCTTAATGAACC3' and 5'CGGTTCATTAAAGGATCCT3' were synthesized using standard cyanoethyl phosphoramidite chemistry, and annealed and kinased to generate a linker molecule which reconstructs the ATG translation initiation codon, includes a stop codon (TAA) for termination of transcription, creates an *Nci*I cohesive end, and adds a *Bam*HI cloning site. Linkers were ligated to the 1.17 kb T4 fragment and then digested with *Bam*HI to generate *Bam*HI cohesive ends. Subsequently, the T4 fragment was inserted into the *Bam*HI cloning site of pAc373. Recombinant plasmids containing a single copy of the truncated CD4 molecule in the correct orientation were identified by restriction mapping. The constructs were then sequenced²⁰ to confirm the sequence at the insertion junctions. Two recombinant plasmids, pAc373/T4_{ex1} and pAc373/T4_{ex2} were characterized in detail. They contained identical 5' ends. The synthetic linker ligated in pAc373/T4_{ex1} to result in a predicted CD4 protein carboxy-terminus of LPTWSTPVH. But in pAc373/T4_{ex2} the 3' linker ligated in the reverse orientation resulting in a cDNA read-through into the polyhedrin gene, such that the predicted carboxy-terminus is LPTWSTPGSFPTQEPKTHSLQGNP. Transfer of the T4_{ex} sequence from the plasmid vector to the AcNPV genome was as described¹³. In this method, cotransfection by calcium phosphate precipitation of 4 µg pAc373/T4_{ex} DNA combined with 1 µg purified AcNPV DNA into *S. frugiperda* cells (SF9) results in homologous recombination between the recombinant sequence of the transfer vector and the polyhedrin gene sequence of AcNPV. Recombinant AcNPV contains an inactivated polyhedrin gene which no longer forms occlusions in infected cells. For plaque purification, 2 × 10⁶ SF9 cells were seeded in 100-mm Petri dishes ~24 h before assay. Tenfold dilutions of viral supernatant were prepared using a final medium: Grace's insect medium (Gibco, NY), Difco TC yeastolate 0.33%, lactalbumin hydrolysate (Gibco) 0.33%, 2 mM supplemental glutamine and 50 µg ml⁻¹ gentamicin containing 10% FCS (Hyclone, UT). Each plate was inoculated with 1 ml virus (10⁻³ to 10⁻⁷ dilution) plus 2 ml final medium. After 2 h incubation, the inoculum was removed and replaced with 10 ml 1.5% Sea Plaque agarose (FMC Bioproducts) in final media. After agarose solidification, plates were transferred to a humid environment for 4–6 d at 27 °C. Plaque assay of the transfection supernatant yielded plaques of distinct morphology; either infected cells which were occlusion-positive (wild-type AcNPV) or occlusion-negative (recombinant T4 virus). Occlusion-negative plaques were identified, selected and further purified. DNA from cells infected with putative T4 recombinant virus was hybridized with a ³²P-labelled T4 cDNA probe to verify the presence of the T4 sequence. Once purified, high titre viral stocks were generated by infecting SF9 cells at an MOI of 1 and culturing at 1 × 10⁶ cells ml⁻¹ for 4 d in final medium. For large-scale production of protein, SF9 cells were grown in 2-litre spinner flasks in final medium. Cells were collected and infected with an MOI of 15 at a concentration of 10 × 10⁶ cells ml⁻¹. Cells were then pelleted, resuspended in media at 1 × 10⁶ ml⁻¹ and cultured for 3 d at 27 °C in spinner flasks. At this time, supernatants were collected by centrifuging cultures to remove cells.

gene. Note that under non-reducing conditions (Fig. 2b, lane 4), T4_{ex1} protein moves faster than under reducing conditions, consistent with the prediction that there are intrachain disulphide bonds in the external segments of CD4 (ref. 14) and also showing the absence of covalent disulphide-linked polymers of T4_{ex1} protein.

Samples of the 50K T4_{ex1} and 51K T4_{ex2} proteins were prepared¹⁵ and subjected to amino-terminal sequencing on an

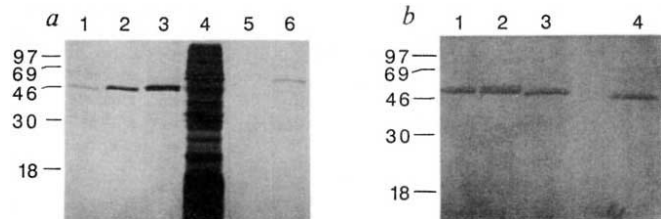


Fig. 2 Purification and characterization of the T4 polypeptides: the major secreted product of T4_{ex} AcNPV-infected SF9 cells. **a**, [³⁵S]methionine metabolic labelling of T4_{ex1}. SF9 cells were infected with either recombinant baculovirus containing the T4_{ex} cDNAs or wild-type AcNPV virus followed by culture in [³⁵S]methionine. Products were autoradiographed after SDS-PAGE. Lanes 1 and 4: whole culture supernatant and cell lysate of T4_{ex1} virus-infected cells respectively; lanes 2 and 3: anti-CD4 precipitated material from T4_{ex1} infected cell supernatant and lysate respectively; lane 5: anti-CD4 precipitated material from wild-type AcNPV-infected SF9 cell supernatants; lane 6: whole culture supernatant of wild-type AcNPV infected cells. Molecular weight markers are phosphorylase B (97.4K), bovine serum albumin (69K), ovalbumin (46K), carbonic anhydrase (30K), lactoglobulin A (18.4K). **b**, SDS-PAGE of T4_{ex1} and T4_{ex2} polypeptides stained with Coomassie blue. Lanes 1 and 2: different preparations of T4_{ex2} polypeptide; lanes 3 and 4: preparations of T4_{ex1} polypeptide. Tracks 1–3 were run under reducing conditions, track 4 under non-reducing conditions.

Methods. **a**, 6 × 10⁵ SF9 cells were seeded per well in 24-well Nunc plates (Interlab, CA) for 2 h at 27 °C and then adherent cells infected with virus at an MOI of 10 in 0.2 ml final medium for 2 h. The inoculum was removed and cells cultured in 0.5 ml fresh medium at 27 °C for 48 h. Adherent cells were washed twice with 0.5 ml Grace's medium lacking serum and methionine, followed by incubation in 0.5 ml in the same medium for 1 h. The adherent cells were washed once and then cultured for 6 h in serum- and methionine-free Grace's medium containing 67 µCi ml⁻¹ [³⁵S]methionine (New England Nuclear, 1,134 Ci mmol⁻¹). Culture supernatants were collected, microfuged for 10 min and dialysed at 4 °C against PBS containing 0.05% sodium azide and 10 mM unlabelled methionine. Cells were dislodged from the wells, washed twice with Grace's medium at 4 °C (by centrifugation in a Sorvall RT600 for 5 min at 1,000 r.p.m.) and finally lysed for 30 min at 4 °C by the addition of RIPA buffer containing 1% Triton X-100, 0.15 M NaCl and protease inhibitors (see below). The lysates were microfuged for 10 min and dialysed at 4 °C as for culture supernatants. Both lysates and culture supernatants were subjected to immunoprecipitation for 16 h at 4 °C with a monoclonal anti-T4 antibody (19Thy5D7) linked to Affigel-10 beads (5 mg monoclonal antibody per ml gel). After immunoprecipitation, the beads were washed twice with lysis buffer and bound material eluted with 0.1 M glycine-HCl buffer, pH 2.0. Eluates and whole samples of lysates or culture supernatants were mixed with SDS sample buffer containing 2-mercaptoethanol, boiled for 5 min and electrophoresed in 12.5% mini-slab gels. Subsequently, the gels were fixed, dried and autoradiographed using Kodak XAR-5 film. **b**, For large scale purification, infected SF9 cell culture supernatants were spun in a Sorvall H-400 rotor at 800 r.p.m. for 6 min at 4 °C. The culture supernatants were treated with protease inhibitors leupeptin, antipain, pepstatin and chymostatin to final concentrations of 0.5 µg ml⁻¹; soybean trypsin inhibitor to 0.02 µg ml⁻¹; and phenyl methyl sulphonyl fluoride to 1.25 mM, and the pH adjusted to 6.8. After subsequent clarification at 8,000 r.p.m. in a Sorvall GSA rotor for 25 min at 4 °C samples were pumped at 4 °C at 30 ml h⁻¹ through a 2 ml pre-clear immunoabsorbent column of 21Thy2D3 monoclonal antibody (anti-T8) coupled to Affigel-10 (Biorad), followed in series by a 7 ml column of anti-CD4 (19Thy5D7) monoclonal antibody coupled to Affigel 10 at a concentration of 7.5 mg monoclonal antibody per ml of gel. The anti-CD4 column was then washed with 30 ml 10 mM Tris-HCl buffer, pH 6.8, followed by 15 ml 0.1 M glycine-HCl buffer, pH 5.0. The bound CD4 polypeptides were eluted by pumping 0.1 M glycine-HCl, pH 2.0, through the washed anti-CD4 column and 0.8 ml fractions of eluate were collected into tubes containing 0.15 M Tris-HCl, pH 7.6. Eluate absorption was monitored at 280 nm with a Uvicord 2 (LKB, Gaithersburg, MD) fitted with an event marker. Fractions of neutralized pH 2.0 eluate containing protein were pooled and concentrated by ultrafiltration in a stirred cell (Amicon, model 3) fitted with a YM-5 membrane. Typically, the yield of purified T4_{ex1} and T4_{ex2} polypeptides was 1 µg ml⁻¹ of infected SF9 culture supernatants. Aliquots containing 1 µg protein concentrate (assuming that 1 absorbance unit represents 1 mg ml⁻¹ at 280 nm) were analysed on 12.5% SDS-polyacrylamide slab gels stained with Coomassie blue. Yields were routinely 1–2 mg secreted T4_{ex1} or T4_{ex2} proteins per litre SF9 cells (1–2 × 10⁶ cells) over a 72 h culture period.

Applied Biosystems model 470A sequencer equipped with an on-line 120A PTH analyser using the 03RPTH program. In both cases the sequence was KKVVLGKKGD. In contrast, the predicted N-terminal sequence of CD4 (based on translation of the cDNA nucleotide sequence) has been suggested to be QGNKVVLGKKGD (ref. 12) or NKVVLGKKGD (ref. 14). The second of these was based on homology with the rat N-terminal sequence, KTVVLGK. Although our empirically

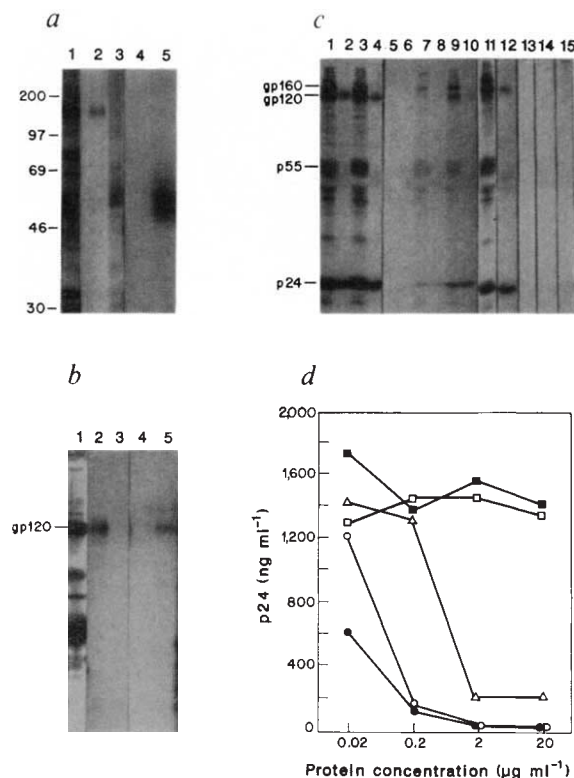
Fig. 3 Soluble CD4 proteins bind to gp120 and inhibit HIV replication. *a*, Coprecipitation of HIV proteins and soluble CD4. SDS-PAGE analysis: lane 1, labelled culture supernatants containing HIV with molecular weight positions indicated; lane 2, labelled HIV proteins incubated with $T_{4_{ex2}}$, immunoprecipitated with OKT4; lane 3, labelled HIV proteins incubated with $T_{4_{ex2}}$ that had been preincubated with OKT4A, immunoprecipitated with OKT4; lane 4, incubation of unlabelled uninfected Molt-3 supernatants with iodinated $T_{4_{ex2}}$, immunoprecipitated with anti-gp120; lane 5, incubation of unlabelled HIV-infected supernatants with iodinated soluble $T_{4_{ex2}}$, immunoprecipitated with anti-gp120. *b*, Inhibition of gp120 binding to cell surface CD4 on SupT1 cells by soluble CD4. gp120 binding to SupT1 cells was determined in the presence of either: $2.5 \mu\text{g ml}^{-1}$ OKT4 (lane 2); $2.5 \mu\text{g ml}^{-1}$ OKT4A (lane 3); $30 \mu\text{g ml}^{-1}$ $T_{4_{ex1}}$ (lane 4); or $30 \mu\text{g ml}^{-1}$ T11 (lane 5). Lane 1 shows labelled HIV proteins from HIV-infected H9 culture supernatants with the position of gp120 indicated. *c*, Effects of soluble T4 proteins on HIV replication in C8166 cells. SDS-PAGE analysis of immunoprecipitates using RV119 patient antiserum of radiolabelled cell lysates (lanes 1, 3, 5, 7, 9, 11, 13, 15) or cell supernatants (lanes 2, 4, 6, 8, 10, 12, 14) after HIV infection in the presence of: no added protein (lanes 1 and 2); ovalbumin, $10 \mu\text{g ml}^{-1}$ (lanes 3 and 4); $T_{4_{ex2}}$ protein, $10 \mu\text{g ml}^{-1}$ (lanes 5 and 6); $T_{4_{ex2}}$ protein, $2 \mu\text{g ml}^{-1}$ (lanes 7 and 8); $T_{4_{ex2}}$ protein, $0.2 \mu\text{g ml}^{-1}$ (lanes 9 and 10); 21Thy2D3, $10 \mu\text{g ml}^{-1}$ (lanes 11 and 12); 19Thy5D7, $0.2 \mu\text{g ml}^{-1}$ (lanes 13 and 14). Lane 15 is a radioimmunoprecipitate of lysates from mock-infected C8166 cells. The positions of the envelope (gp160 and gp120) and gag proteins (p55 and p24) are noted. *d*, Effects of soluble CD4 proteins on HIV replication in H9 cells. The amount of p24 protein in cell supernatants is plotted against the concentration of added ovalbumin ($\square$), soluble T11 produced in baculovirus ($\blacksquare$), $T_{4_{ex1}}$ protein ($\circ$), $T_{4_{ex2}}$ protein ($\bullet$) or 19Thy5D7 (Δ).

Methods. *a*, Culture supernatants were collected from 1×10^7 Molt-3 lymphocytes stably infected with the HIV/HTLV IIIB that were metabolically labelled overnight with $100 \mu\text{Ci ml}^{-1}$ [^{35}S]cysteine (NEN) in total volume of 1.5 ml. NP-40 was added to 0.5% final concentration and the supernatants were incubated with $10 \mu\text{g T}_{4_{ex2}}$ for 1 h at 37°C with or without preincubation of the soluble CD4 with $5 \mu\text{g ml}^{-1}$ of OKT4A (Ortho Pharmaceutical). The samples were then immunoprecipitated with the monoclonal antibody OKT4 (Ortho) and run on SDS-polyacrylamide gels¹¹. In addition, 1.5 ml of unlabelled culture supernatants were collected from either 1×10^6 uninfected Molt-3 lymphocytes or HIV-infected Molt-3 lymphocytes and incubated for 1 h at 37°C with ^{125}I -labelled $T_{4_{ex2}}$ ($0.9 \mu\text{g}$, specific activity 1.4×10^6 c.p.m. μg^{-1}) that had been radioiodinated by Bolton-Hunter reagent (NEN). The samples were then immunoprecipitated using a goat anti-gp120 antiserum²¹. *b*, 1×10^7 H9 lymphocytes stably infected with the HIV strain IIIB were metabolically labelled with [^{35}S]cysteine overnight in 1.5 ml. The supernatants containing labelled HIV proteins were incubated for 1 h at 37°C with 1×10^7 SupT1 lymphocytes which had been preincubated for 30 min at 37°C with various proteins. The SupT1 cells were spun, washed with PBS, lysed with 0.75 ml lysis buffer and the gp120 bound to the SupT1 cells was immunoprecipitated with RV119 AIDS patient antiserum²¹. *c*, 2×10^5 C8166 cells were infected with 1,000 TCID₅₀ units of the H9IIIB isolate of HIV 7 d previously. Proteins were added at the time of infection and maintained at the concentrations indicated throughout the course of the experiment. *d*, H9 lymphocytes were infected as described in *c*. On day 9 following infection, supernatants were collected and assayed for viral p24 gag protein by radioimmunoassay.

derived sequence is consistent with the positioning suggested in ref. 14 and with that of CD4 in mouse and sheep¹⁶, the K at position 1 is at variance with the amino acid predicted from the cDNA sequence. We sequenced the DNA of the pAc373/ $T_{4_{ex1}}$ and pAc373/ $T_{4_{ex2}}$ inserts and found the codon for the amino-terminal residue to be AAG rather than AAC (ref. 12) without other differences. This may be a mutation resulting from cloning into pAc373 but this seems unlikely as a lysine residue is found at the N-terminus of the homologous rat CD4 sequence. We conclude that the amino terminus of mature human CD4 begins KKV and that $T_{4_{ex1}}$ and $T_{4_{ex2}}$ are CD4-derived polypeptides. These data show that the baculovirus expression system can enzymatically cleave the signal peptide from the $T_{4_{ex}}$ polypeptide precursor, allowing it to be secreted.

To determine whether these soluble CD4 proteins could bind the HIV gp120 exterior glycoprotein, we performed reciprocal coprecipitation experiments (Fig. 3*a*). Metabolically labelled gp120 protein derived from HIV virions was incubated with unlabelled purified $T_{4_{ex2}}$ protein with or without two monoclonal antibodies directed against distinct epitopes of the CD4 protein (OKT4 and OKT4A). OKT4A, like 19Thy5D7, inhibits binding of gp120 to the CD4 molecule⁶ but OKT4 does not. Without OKT4A preincubation, gp120 protein was coprecipitated by OKT4 antibody (Fig. 3*a*, lane 2). In contrast, preincubation with OKT4A antibody inhibits gp120 co-precipitation (Fig. 3*a*, lane 3). Thus gp120 binds to the $T_{4_{ex2}}$ protein, and this binding is inhibited by OKT4A. OKT4 does not precipitate gp120 in the absence of $T_{4_{ex2}}$ protein.

In a second experiment, unlabelled HIV virions were incubated with radioiodinated $T_{4_{ex2}}$ protein, and the mixture immunoprecipitated with a goat antiserum raised against purified HIV gp120 protein. The iodinated $T_{4_{ex2}}$ protein was coprecipitated by the anti-gp120 serum only when HIV virions were present (Fig. 3*a*, compare lanes 5 and 4), suggesting that the $T_{4_{ex2}}$ protein binds to an HIV virion component.



To examine whether the $T_{4_{ex1}}$ protein inhibits the binding of gp120 protein to CD4⁺ lymphocytes, metabolically labelled gp120 protein from the supernatants of virus-infected cells was preincubated with $T_{4_{ex1}}$ protein or, as a control, the extracellular T11 protein segment made in the baculovirus system (our unpublished results) and the mixtures incubated with CD4⁺ SupT1 lymphocytes. Cells were lysed and the solubilized material immunoprecipitated with antiserum from an AIDS patient which contained anti-gp120 reactivity. To control for specificity, OKT4 and OKT4A monoclonal antibodies were added to the SupT1 cells before the labelled protein. As expected, OKT4A but not OKT4 inhibited binding of the labelled gp120 protein to the SupT1 cells (Fig. 3*b*, compare lanes 3 and 2). The $T_{4_{ex1}}$ protein significantly inhibited binding of labelled gp120 to the surface of SupT1 cells (Fig. 3*b*, lane 4) but no inhibition was observed using up to $30 \mu\text{g ml}^{-1}$ of the control protein (Fig. 3*b*, lane 5), whereas inhibition of gp120 binding was seen at $0.5 \mu\text{g ml}^{-1}$ $T_{4_{ex1}}$ protein (data not shown). Comparable inhibition was found for $T_{4_{ex2}}$.

We examined whether induction of syncytia by HIV envelope, which depends on binding of the gp120 glycoprotein to the CD4 molecule followed by events involved in membrane fusion, could be inhibited by $T_{4_{ex2}}$. Jurkat cells (10^4) infected with the HXBc2 strain of HIV or H9 cells infected with HTLV-III_B were cocultivated with 5×10^5 CD4⁺ SupT1 lymphocytes in 1 ml in the presence or absence of the $T_{4_{ex2}}$ protein and syncytium formation was assayed at 10 h. Addition of control proteins (ovalbumin or an anti-CD8 monoclonal antibody, 21Thy2D3), to the cocultivated cells even to a concentration of $50 \mu\text{g ml}^{-1}$ had no effect on the formation of syncytia ($>1,000$ syncytia per ml), but addition of as little as $2 \mu\text{g ml}^{-1}$ $T_{4_{ex2}}$ or anti-T4 (19Thy5D7) completely inhibited it (<50 syncytia per ml). Both $T_{4_{ex1}}$ and $T_{4_{ex2}}$ proteins also inhibited the induction of syncytia when CHO cells constitutively expressing HIV envelope were cocultivated with SupT1 lymphocytes, whereas recombinant secreted

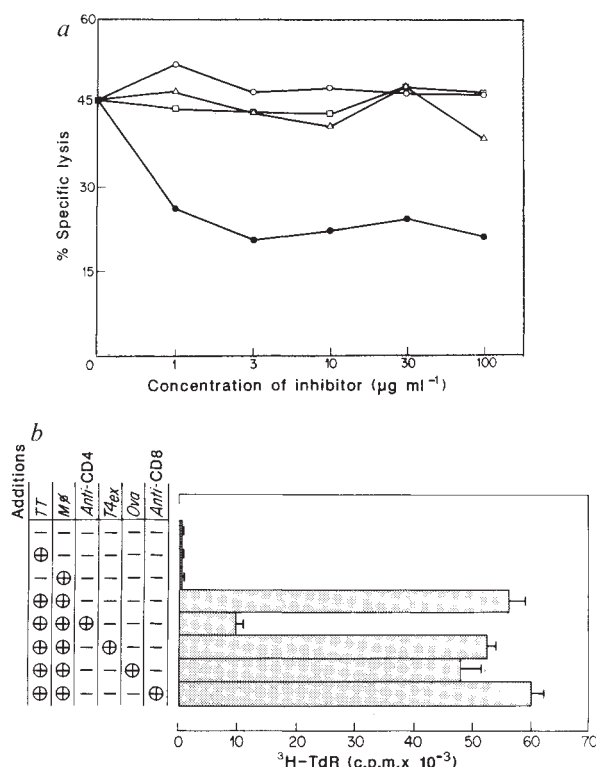


Fig. 4 T4_{ex} proteins do not inhibit class II MHC recognition events. *a*, The ability of secreted CD4 protein to inhibit cytotoxicity by a T4+T8-alloreactive cytotoxic clone was measured in the presence of T4_{ex2} (Δ), BSA (□), anti-CD4 (●) or anti-CALLA (○). *b*, Effects of secreted T4_{ex1} protein on proliferation of a tetanus toxoid specific, class II MHC restricted clone, CTT7. Plus (+) sign indicates the presence of a given additive.

Methods. *a*, The T4+ cytotoxic clone AA8 is specific for class II MHC gene products on the allogeneic EBV transformed B cell line Laz 509 (ref. 18). ⁵¹Cr-labelled Laz 509 cells were preincubated with secreted T4_{ex2} or a control protein (BSA) for 30 min at 4 °C before addition of effector cells. In other wells, anti-CD4 (19Thy5D7) and anti-CALLA J5 (gift from Jerome Ritz) antibodies were used as inhibitors. After preincubation of targets (3,000 cells per well) with secreted CD4 and control inhibitors, lysis was measured in a standard 4 h ⁵¹Cr release assay at an E/T ratio of 30:1. Results shown are the mean of quadruplicate samples where standard deviations are <10%. *b*, For proliferative studies, 50,000 cells per well of the tetanus toxoid specific clone CTT7, derived by standard cloning strategies¹, were cultured in 10% FCS/RPMI 1640 supplemented with 1% penicillin-streptomycin and 2% glutamine alone, with 5 µg ml⁻¹ of tetanus toxoid (Massachusetts Department of Public Health) or the combination of tetanus toxoid (TT) and 10% autologous macrophages (Mφ) in the presence or absence of a final concentration of 40 µg ml⁻¹ ovalbumin (ova), anti-CD4, anti-CD8 or twice immunoabsorbed T4_{ex1} for 24 h at 37 °C in a humid atmosphere with 6% CO₂. Subsequently, wells were pulsed with 1 µCi per well of ³H-TdR. Cells were collected at 48 h.

gave significant inhibition of viral protein expression and virus production. T4_{ex1} and T4_{ex2} proteins decreased HIV p24 protein expression at 0.2 µg ml⁻¹ concentrations.

Because CD4 facilitates the activation of class II specific CTL and Ia-restricted helper T lymphocytes^{1,17}, we examined whether soluble CD4 abrogates physiological responses of CD4⁺ T lymphocytes. First we examined the effect of T4_{ex2} on class II MHC-directed killing mediated by the CD4⁺ cytotoxic clone AA8 (ref. 18) (Fig. 4*a*). As shown, T4_{ex2}, like the control protein BSA and the anti-CALLA monoclonal antibody, failed to inhibit CTL effector function even at concentrations as high as 100 µg ml⁻¹. In contrast, as little as 1–3 µg ml⁻¹ specific anti-CD4 monoclonal antibody reduces cytotoxicity by >50%. T4_{ex1} was also without effect on cell lysis (data not shown). We also examined the effect of T4_{ex1} on proliferation of the CD4⁺ tetanus toxoid-specific, class II MHC restricted helper T cell clone, CTT7. The CTT7 clone proliferates only in the presence of tetanus toxoid and the autologous antigen-presenting cell. At 40 µg ml⁻¹, the anti-CD4 (19Thy5D7) monoclonal antibody inhibited ³H-TdR incorporation by >80%, consistent with the important function of CD4 in helper T-cell responses. In contrast, equivalent amounts of T4_{ex1}, ovalbumin or anti-CD8 monoclonal antibody have no effect. The basis for this difference remains to be resolved. One possibility is that the affinity of CD4 for gp120 is substantially higher than CD4 for its native ligand (presumably class II MHC). In addition, because CD4 is only one of several elements (including LFA-1 and T11), that facilitates cell-cell interactions between CTL and targets or inducer T cells and antigen-presenting cells¹⁹, partial abrogation of the CD4 function with T4_{ex1} protein may leave the T cell activation process uninhibited. Alternatively, truncation of the CD4 protein may result in loss or modification of regions (in the carboxy-terminal region or elsewhere) of the CD4 molecule involved in interaction with class II MHC. This seems unlikely as each of three anti-CD4 monoclonal antibodies (19Thy5D7, 18T3A9, and OKT4A) tested, all of which block class II MHC recognition events, reacts with T4_{ex1} and T4_{ex2} protein. It is also possible that anti-CD4 antibodies inhibit T cell responses by means other than blockade of the external domain of CD4. The availability of soluble CD4 proteins allows the development of assays to detect drugs which could interfere with gp120-CD4 interactions, and experiments to determine whether the class II-binding site of CD4 is distinct from the gp120 binding site. T4_{ex} proteins or fragments derived from them could be clinically useful in inhibiting gp120 binding to membrane-bound CD4 on T lymphocytes, monocytes or brain cells, without interfering with the normal physiological function of surface CD4 on healthy cells. Finally, this type of eukaryotic expression system could also facilitate structure-function analyses of other receptor proteins.

Received 8 October; accepted 23 November 1987.

- Meuer, S., Schlossman, S. F. & Reinherz, E. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4395–4399 (1982).
- Biddison, W., Rao, P., Thalle, M., Goldstein, G. & Shaw, S. J. *exp. Med.* **156**, 1065–1076 (1982).
- Gay, D. *et al. Nature* **328**, 626–629 (1987).
- Klatzmann, D. *et al. Science* **225**, 59–63 (1984).
- Dagleish, A. *et al. Nature* **312**, 763–766 (1984).
- Sattentau, Q., Dagleish, A., Weiss, R. & Beverley, P. C. L. *Science* **234**, 1120–1123 (1986).
- McDougal, J. S. *et al. J. Immun.* **137**, 2937–2944 (1986).
- McDougal, J. S. *et al. Science* **231**, 382–385 (1986).
- Maddon, P. *et al. Cell* **47**, 333–348 (1986).
- Sodroski, J., Goh, W. C., Rosen, C., Campbell, K. & Haseltine, W. A. *Nature* **322**, 470–474 (1986).
- Lifson, J. *et al. Nature* **323**, 725–728 (1986).
- Maddon, P. *et al. Cell* **42**, 93–104 (1985).
- Smith, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 8404–8408 (1985).
- Clark, S., Jefferies, W., Barclay, A. N., Gagnon, J. & Williams, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1649–1653 (1987).
- Matsudaira, P. *J. biol. Chem.* **262**, 10035–10038 (1987).
- Classon, B. *et al. Immunogenetics* **23**, 129–132 (1986).
- Meuer, S. C. *et al. Science* **222**, 1239–1242 (1983).
- Acuto, O., Hussey, R. E. & Reinherz, E. L. *J. exp. Med.* **162**, 1387–1392 (1985).
- Krensky, A., Robbins, E., Springer, T. A. & Burakoff, S. J. *Immun.* **132**, 2180–2184 (1984).
- Biggin, M., Gibson, T. & Hong, G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3963–3965 (1983).
- Kowalski, M. *et al. Science* **237**, 1351–1355 (1987).

T11 protein made in a baculovirus expression system did not (data not shown). Pretreatment of the envelope-expressing cells (H9/HTLV-III_B) with T4_{ex2} also inhibited syncytium formation but preincubation of the target SupT1 cells gave only slight inhibition. Thus, T4_{ex2} protein appears to inhibit syncytium formation by interaction with the envelope-expressing cells.

To determine whether soluble CD4 inhibits virus entry and replication, H9 and C8166 lymphocytes were infected with the H9II_B strain of HIV in the presence of T4_{ex1} or T4_{ex2} proteins. Control proteins in this experiment were ovalbumin, a recombinant T11 protein made in baculovirus, and the monoclonal antibodies anti-CD4 (19Thy5D7) and anti-CD8 (21Thy2D3). Radioimmunoprecipitations of labelled cell lysates and supernatants using antiserum from an AIDS patient were used to assess replication. The amount of viral p24 antigen present in the cell supernatant was measured throughout the course of infection. Figure 3*c* and *d* shows that ovalbumin, recombinant T11 and 21Thy2D3 proteins had no effect on virus replication, even at 20 µg ml⁻¹ but T4_{ex1}, T4_{ex2} and anti-CD4 (19Thy5D7)

A soluble form of CD4 (T4) protein inhibits AIDS virus infection

Keith C. Deen*, J. Steven McDougal†, Richard Inacker*, Gail Folena-Wasserman*, Jim Arthos‡, Jonathan Rosenberg*, Paul Jay Maddon§, Richard Axel||¶ & Raymond W. Sweet*

* Smith Kline and French Laboratories, King of Prussia, Pennsylvania 19406, USA

† Centers for Disease Control, Atlanta, Georgia 30333, USA

‡ Department of Human Genetics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

§ Department of Biochemistry and Molecular Biophysics and

|| Howard Hughes Medical Institute, Columbia University, 722 West 168th St, New York, New York 10032, USA

CD4 (T4) is a glycoprotein of relative molecular mass 55,000 (M_r 55K) on the surface of T lymphocytes which is thought to interact with class II MHC (major histocompatibility complex) molecules, mediating efficient association of helper T cells with antigen-bearing targets¹⁻³. The CD4 protein is also the receptor for HIV, a T-lymphotropic RNA virus responsible for the human acquired immune deficiency syndrome (AIDS) (refs 4-7). To define the mechanisms of interaction of CD4 with the surface of antigen-presenting cells and with HIV, we have isolated the CD4 gene and expressed this gene in several different cellular environments^{7,8}. Here we describe an efficient expression system in which a recombinant, soluble form of CD4 (sCD4) is secreted into tissue culture supernatants. This sCD4 retains the structural and biological properties of CD4 on the cell surface, binds to the envelope glycoprotein (gp110) of HIV and inhibits the binding of virus to CD4⁺ lymphocytes, resulting in a striking inhibition of virus infectivity.

The CD4 molecule is comprised of an N-terminal hydrophobic signal sequence, four extracellular domains with homology to immunoglobulin variable and joining regions, a hydrophobic transmembrane domain, and a highly charged cytoplasmic segment^{8,9}. To obtain large quantities of the extracellular segment of CD4, we introduced a termination codon into the CD4 complementary DNA⁸ at the extracellular boundary of the transmembrane domain, expecting that the truncated protein would be secreted in a soluble form. This sCD4 cDNA was inserted between the SV40 early promoter and the bovine growth hormone (BGH) polyadenylation site (Fig. 1a). The sCD4 expression cassette was linked to a mouse dihydrofolate reductase (*dhfr*) expression cassette¹⁰ to permit co-amplification of the sCD4 gene on exposure to increasing concentrations of the *dhfr* inhibitor methotrexate (MTX). This vector, psT4DHFR (Fig. 1a), was transfected into the *dhfr*-deficient CHO cell line, DXB11 (ref. 11), and transformants were selected and exposed to progressively increasing concentrations of methotrexate¹². Supernatants from clones were monitored for the expression of sCD4 by one of three independent assays: (1) immunoprecipitation from ³⁵S-labelled cultures with CD4-specific monoclonal antibodies; (2) Western blot analysis using a rabbit anti-CD4 polyclonal antibody developed against a denatured CD4 protein produced in bacteria; and (3) competition enzyme-linked immunosorbent assay (ELISA) performed with immobilized sCD4.

³⁵S-methionine- and ³⁵S-cysteine-labelled sCD4 from culture supernatants was specifically immunoprecipitated by each of six anti-CD4 monoclonal antibodies but not by control antibodies (Fig. 1b). In contrast, when culture supernatants were analysed under denaturing conditions by Western blot analysis, sCD4 was not recognized by any of the six anti-CD4 monoclonal

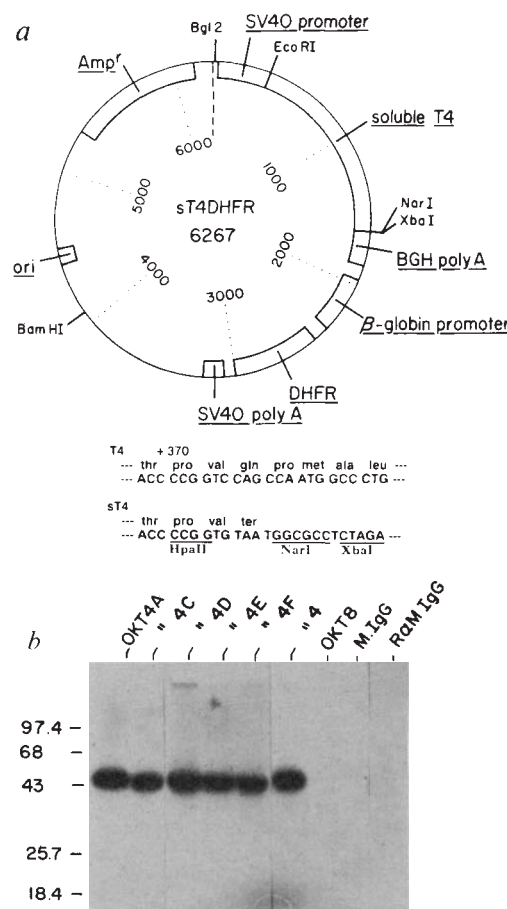


Fig. 1 a, Plasmid psT4DHFR is a pUC18 derivative containing base pairs (bp) 1-1,257 of the CD4 cDNA clone pT4B⁸ which encodes the leader and extracellular segment of CD4. This sCD4 cDNA is inserted between an SV40 early promoter and a synthetic linker containing a TAA termination codon (inset) followed by the polyadenylation region of the bovine growth hormone gene. The sCD4 expression cassette is linked to a mouse *dhfr* expression cassette consisting of the β -globin promoter, mouse *dhfr* coding sequence and the SV40 polyadenylation region. b, Epitope characterization of sCD4. Cell lines were metabolically labelled with ³⁵S-methionine and cysteine and the supernatants were immunoprecipitated with monoclonal antibodies to CD4 (OKT4, T4, T4C-F), CD8, and control antibodies (mouse IgG (M. IgG) and rabbit anti-mouse IgG (RαM IgG)). Regular molecular masses (M_r) are given in thousands.

Methods. a, Plasmid psT4DHFR was constructed in pUC18 and contains the following sequence elements isolated from previously reported plasmids: SV40 early promoter¹⁰, SV40 poly(A) early region¹⁰, bovine growth hormone poly(A) region¹⁰, mouse β -globin promoter¹⁴, mouse *dhfr* coding region¹⁵ and bp 1-1,257 of the CD4 cDNA clone pT4B⁸. The full-length CD4 cDNA was truncated at a *Hpa*II site (bp 1,252) and ligated to a synthetic linker (inset) which restored bp 1,253-1,257 of the CD4 coding sequence and placed a TAA termination codon after bp 1,257. DXB-11 cells, a *dhfr*⁻ CHO cell line¹¹, were transfected by calcium phosphate precipitation¹⁶ with 10 or 30 μ g of psT4DHFR one day after seeding. Transformants were selected in nucleoside-free medium and the transforming sequences were amplified through rounds of increasing MTX concentrations¹². b, Cultures containing 1×10^5 cells per 60 mm culture dish were labelled for 16 h at 37 °C in 1.5 ml methionine and cysteine free F12 medium containing ITS (Collaborative), glutamine, 170 μ Ci ml⁻¹ ³⁵S-methionine and 30 μ Ci ml⁻¹ ³⁵S-cysteine (ICN). Clarified media was precleared, incubated with 5 μ g OKT4, T4, T4C-T4F, OKT8 (P. Rao, Ortho), mouse IgG (Cooper), or rabbit anti-mouse IgG (Cooper) for 30 min at 4 °C, then incubated with protein A-Sepharose beads. The beads were washed twice, boiled for 5 min in 20 μ l sample buffer and analysed by electrophoresis through a 12.5% SDS-polyacrylamide gel under reducing conditions.

¶ To whom correspondence should be addressed.

Fig. 2 a, SDS-PAGE profile of purified sCD4. Relative molecular masses (M_r) are given in kilodaltons. **b**, Coprecipitation of HIV gp110 with sCD4. Supernatant containing sCD4 and control supernatant from untransformed DXB-11 cells were adsorbed to Sepharose beads coated with OKT4, control antibody MOPC141 (isotype matched to OKT4), or human anti-HIV IgG. The beads were washed, mixed with a lysate of 35 S-methionine-labelled HIV, washed, and bound material was eluted and subjected to SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Lane 1, eluate from OKT4 beads incubated with sCD4; lane 2, eluate from OKT4 beads incubated with control supernatant; lane 3, eluate from MOPC141 beads incubated with sCD4; lane 4, eluate from anti-HIV beads; lane 5, eluate from non-immune IgG beads. Relative molecular masses are given.

Methods. **a**, Purified sCD4 (3 μ g) was electrophoresed through a 12.5% polyacrylamide gel and visualized by silver stain. **b**, A lysate of extracellular HIV was internally labelled with 35 S-methionine and 35 S-cysteine as previously described⁶. Sepharose beads coupled to OKT4 monoclonal antibody, MOPC141 paraprotein (isotype matched to OKT4), human anti-HIV IgG and non-immune human IgG were prepared as previously described⁶. OKT4 beads and MOPC141 beads (10 μ l) were incubated for 2 h at 4 °C with ~ 5 μ g of sCD4, or control supernatant from untransformed DXB-11 cells, then washed once with lysing buffer. The 35 S-labelled HIV (preadsorbed) was added and the beads were incubated for 3 h at 4 °C. Washing, elution of bound material and SDS-PAGE under reducing conditions were performed as previously described⁶.

antibodies (which recognize only native protein), but was readily detected by rabbit anti-CD4 polyclonal serum raised against a denatured form of CD4 produced in bacteria (not shown). Several clones synthesize >3 pg of sCD4 per cell in 24 hours, giving ~ 40 mg of sCD4 per litre in high-density suspension cultures over 4 days. We have purified large quantities of sCD4 to $>95\%$ purity (Fig. 2a), and this purified protein shows the same pattern of antibody reactivity as described for sCD4 in culture supernates. These data suggest that sCD4 maintains a configuration analogous to the native extracellular domain of CD4 on the lymphocyte surface.

We next examined whether sCD4 associates with HIV envelope glycoprotein gp110. Purified sCD4 (~ 5 μ g) was adsorbed to Sepharose beads coated with OKT4 or control antibody. OKT4 associates with the CD4 molecule without inhibiting the ability of CD4 to interact with virus⁴⁻⁷. The beads were mixed with a lysate of 35 S-methionine-labelled HIV. Only the 110K envelope glycoprotein is coprecipitated by the complex of sCD4 with OKT4 (Fig. 2b, lane 1); no viral proteins are precipitated by OKT4 beads without sCD4 or in the presence of control supernatants from the untransformed CHO cells (Fig. 2b, lane 2). Furthermore, no viral protein is precipitated if Sepharose beads coated with control mouse immunoglobulin (isotype matched to OKT4) are incubated with sCD4 (Fig. 2b, lane 3). Thus the soluble CD4 protein, without other T-cell surface components can specifically associate with the envelope glycoprotein of the AIDS virus.

To examine whether sCD4 abolishes HIV binding to the surface of CD4⁺ cells, CEM T cells (which carry CD4) were exposed to HIV in the presence or absence of sCD4. After viral

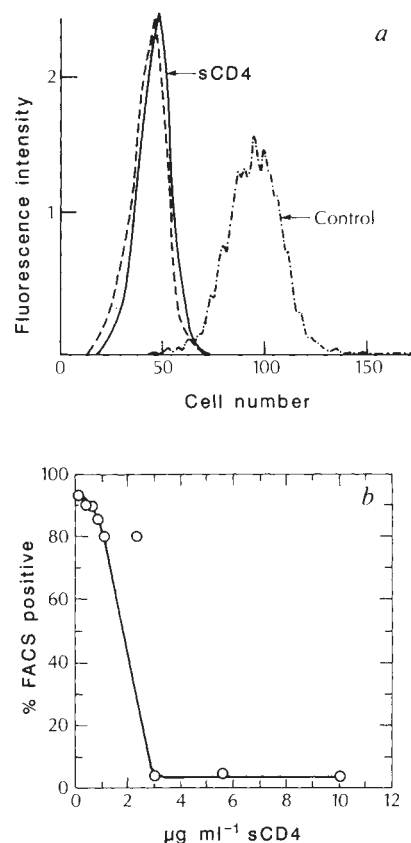


Fig. 3 a, sCD4 inhibits HIV binding to CD4⁺ CEM T cells. Cells were incubated with buffer (---), HIV preincubated with sCD4 (—), or with HIV preincubated with concentrate control supernatant from untransformed DXB-11 (— · —), washed, exposed to fluorescein-conjugated anti-HIV antibody, and analysed by cytometry as previously described¹¹. A fluorescence histogram (cell number against fluorescence intensity) is shown. **b**, Plot of percentage of positive cells by flow cytometry versus concentration of sCD4.

adsorption, the cells were washed, exposed to fluorescein-conjugated anti-HIV antibody, and analysed by flow cytometry (Fig. 3a)^{6,7}. HIV binds efficiently to CEM cells, but if the virus is preincubated with sCD4 the binding is abolished; 3 ng of purified sCD4 is sufficient to inhibit the binding of 100 ng of viral protein (Fig. 3b). We estimate that if the envelope glycoprotein comprises 5% of the total viral protein, then a 2:1 molar ratio of CD4 to gp120 can completely inhibit HIV binding to CD4⁺ cells.

Finally, we have examined the ability of soluble CD4 to inhibit the infection of CD4⁺ cells by HIV. Phytohaemagglutinin (PHA)-stimulated human lymphocytes were exposed to serial ten-fold dilutions of an HIV inoculum in the presence or absence of sCD4, washed and plated in microculture. The frequency of infected culture was determined using an immunoassay 4, 8 and 12 days after exposure to virus¹³ (Fig. 4a). We define the infectious virus titre, ID-50, as the reciprocal of the dilution required to infect 50% of the exposed cell cultures at day 12 (Fig. 4b). In the absence of sCD4, the ID-50 observed with our viral inoculum is $\sim 10^5$. However, in the presence of 8 μ g ml⁻¹ purified sCD4, the infectivity is diminished to an ID-50 of $10^{1.5}$ (Fig. 4b). This dramatic reduction in infectivity is observed throughout the course of infection. As a control for nonspecific inhibition or toxic effects, sCD4 was added to cultures 18 h after the initial exposure to virus. These cultures show only a 1 log inhibition in the ID-50 (Fig. 4b), probably resulting from inhibition of virus spread after the initial inoculation. Thus, the inhibition of virus infection by sCD4 is likely to be a result of

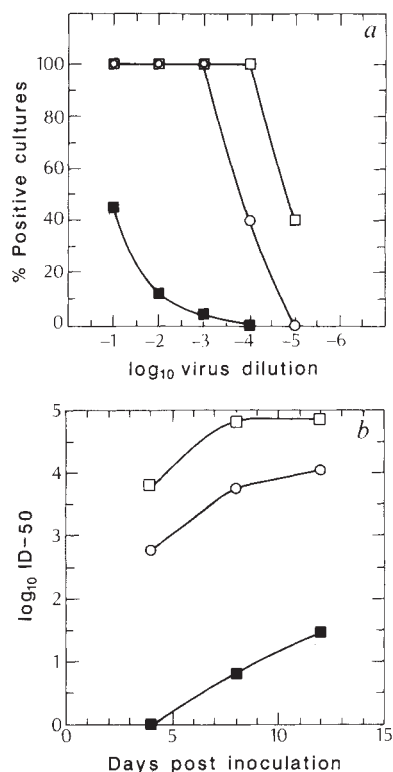


Fig. 4 Inhibition of HIV infectivity by sCD4. *a*, Plot of percentage of cultures positive for HIV at day 8 versus dilution of virus inoculum; *b*, Plot of ID-50 (reciprocal of virus dilution at which 50% of cultures are positive) at days 4, 8 and 12. Infectivity titration of an HIV inoculum (ID-50 assay) was performed as previously described¹⁷. Briefly, serial 10-fold dilutions of virus inoculum are incubated with indicator cells (PHA-stimulated human lymphocytes) for 18 h. The cells are then washed and plated in micro-culture (1×10^5 cells per culture, 10 cultures per dilution). At days 4, 8 and 12, supernatants are tested for HIV by the antigen capture assay¹³. ID-50 titrations were performed in media containing $8.6 \mu\text{g ml}^{-1}$ sCD4 which was added to the HIV dilution 30 min before inoculation of cells and maintained in the culture media throughout the experiment (■—■), or in media containing sCD4 introduced after the initial 18 h inoculum (○—○) (delayed addition control), or in media without sCD4 (□—□), control).

the specific association of sCD4 with gp110, inhibiting the interaction of virus with CD4⁺ cells. We estimate that our viral preparations of 10^4 infectious particles per ml actually contain 10^8 particles per ml. If each particle contains 1,000 envelope glycoproteins, the 3.5 log inhibition we observe is obtained at a ratio of 1,000 CD4 molecules per molecule of envelope glycoprotein.

The ability of sCD4 to bind gp110 and inhibit viral infection *in vitro* immediately suggests the potential use of sCD4 as an antiviral agent in AIDS patients. Although significant variance exists among the different HIV isolates, each appears to use CD4 as a cellular receptor. Thus, sCD4 is likely to be a universal inhibitor of viral infection. Although sCD4 is extremely effective *in vitro*, its use in man will depend on the pharmacokinetics of sCD4, its immunogenicity, its effects on the cellular immune response and the clinical significance of inhibiting viral spread in infected individuals. Whatever the efficacy of sCD4 as a therapeutic in AIDS, this reagent should allow a more precise understanding of the structural basis for the association of CD4 with the HIV envelope glycoprotein and with the surface of antigen-presenting cells.

We thank S. P. Cort, M. S. Kennedy, A. Kandil, C. Silverman and A. Truneh for assistance during the project, M. Rosenberg

for support and encouragement, Ortho Pharmaceuticals for providing OKT4 and Brian Soda for preparation of the figures. This work was supported by the Howard Hughes Medical Institute (R.A.) and by a grant from the National Institutes of Health (P.J.M. and R.A.).

Received 28th October; accepted 26 November 1987.

1. Gay, D. *et al.* *Nature* **328**, 626–629 (1987).
2. Sleckman, B. D. *et al.* *Nature* **328**, 351–353 (1987).
3. Thomas, Y., Rogozinski, L. & Chess, L. *Immun. Rev.* **74**, 113–128 (1985).
4. Dalgleish, A. G. *et al.* *Nature* **312**, 763–766 (1984).
5. Klatzman, D. *et al.* *Nature* **312**, 767–768 (1984).
6. McDougal, J. S. *et al.* *Science* **231**, 382–385 (1986).
7. Maddon, P. J. *et al.* *Cell* **47**, 333–348 (1986).
8. Maddon, P. J. *et al.* *Cell* **42**, 93–104 (1985).
9. Maddon, P. J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
10. Pfarr, D. S., Sathe, G. & Reff, M. F. *DNA* **4**, 461–467 (1985).
11. Umlaud, G., Kas, E., Carothers, A. M. & Chasin, L. A. *Cell* **33**, 405–412 (1983).
12. Kaufman, R. J., Sharp, P. A. & Latt, S. A. *Molec. cell. Biol.* **3**, 699–711 (1983).
13. McDougal, J. S. *et al.* *J. immun. Meth.* **76**, 171–183 (1985).
14. Berg, P. E. *et al.* *Molec. cell. Biol.* **3**, 1246–1254 (1983).
15. Subramani, S., Mulligan, R. & Berg, P. *Molec. cell. Biol.* **1**, 854–864 (1981).
16. Wigler, M. *et al.* *Cell* **11**, 223–232 (1977).
17. McDougal, J. S. *et al.* *J. Immun.* **135**, 3151–3162 (1985).

Soluble CD4 molecules neutralize human immunodeficiency virus type 1

André Traunecker, Wolfgang Lüke* & Klaus Karjalainen

Basel Institute for Immunology, Grenzacherstrasse 487, CH-4058 Basel, Switzerland

* Deutsches Primatenzentrum, Abteilung Virologie und Immunologie, Kellnerweg 4, Göttingen 3400, FRG

Human immunodeficiency virus (HIV) infection can bring about total collapse of the immune system^{1,2} by infecting helper T lymphocytes which express CD4, the molecule which mediates interaction between the cell surface and viral envelope glycoprotein gp120 (refs 3–10). HIV apparently escapes the effects of neutralizing antibodies *in vivo* by generating new variants which must still interact with CD4 to maintain a cycle of infection^{11–14}. One route to block HIV infection, therefore, could use solubilized CD4 protein to inhibit attachment of the virus to its target cell. We have used recombinant DNA techniques to generate soluble forms of CD4, and show here that these are potent inhibitors of HIV infection *in vitro*.

Two different chimaeric genes based on immunoglobulin-expression systems were made (Fig. 1). The construct HT4-Y1 codes for the N-terminal immunoglobulin-like region and three following domains of CD4 which comprise the whole extracellular portion; the other, HT4-X6, codes for the immunoglobulin-like and second domains only^{15,16}. In both cases the C-terminal part of the protein consists of the immunoglobulin κ light-chain constant region.

The introduction of these constructs into myeloma cells (Fig. 2) resulted in the secretion of the two chimaeric proteins of predicted apparent molecular weight: HT4-Y1 and HT4-X6 transformants produced molecules of relative molecular mass (M_r) 60,000 (60K) and 30K respectively, as detected by Western blotting with anti- κ antibodies (Fig. 2a). Furthermore, immunoprecipitations of the supernatants with two different monoclonal anti-CD4 antibodies (Fig. 2b) showed that these secreted proteins retained at least some of their original conformation.

In addition, the soluble CD4 proteins could bind at low concentration to purified HIV-envelope protein gp120 in solid-phase assays, consistent with a high-affinity interaction¹⁰ (Fig. 3). The lower intensity of the signal generated by the 'complete' CD4 (Y124) compared to the truncated one (X6-10) is probably due to a concentration difference in the supernatants (20 and

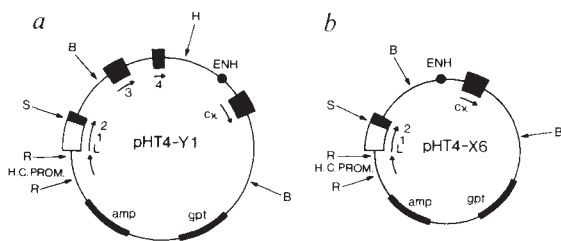


Fig. 1 Chimaeric constructs used to solubilize CD4. *a*, pHT4-Y1, encoding the whole extracellular CD4 region; L (leader) 1, 2, 3 and 4 refer to different sequences encoding protein domains¹⁸. *b*, pHT4-X6, encoding a reduced form of CD4. The arrows under the exons show the transcriptional orientations of the genes. B, *Bam*HI; H, *Hind*III; S, *Sac*I; R, *Eco*RI.

Methods. Exons coding for different parts of CD4 were isolated from a λ EMBL-3 library made from partial *Sau*3A-digested human placental DNA by exploiting exon-specific oligonucleotide probes. The genomic part of CD4 (closed boxes) in pHT4-Y1 was first subcloned into pUC19 as a 2.7 kilobase (kb)-*Sac*I/*Nsi*I restriction fragment. The 5' end of the CD4 was generated by cloning the *Eco*RI/*Sac*I fragment (which encodes leader and V-like sequences and part of the second domain up to amino acid position 152) from CD4 complementary DNA (ref. 15) (open box) into this intermediate pUC19 plasmid. Then the *Eco*RI/*Hind*III fragment containing the CD4-coding information was cloned into the vector p29.7cc3 (ref. 17) (pHT4-Y1-p⁻). Finally, the expression vector (pHT4-Y1) was completed by inserting the immunoglobulin heavy chain (HC) promoter as an *Eco*RI fragment. The vector pHT4-X6 was derived from pHT4-Y1 by replacing the 4.4 kb *Bam*HI fragment with a 2.8 kb *Bam*HI fragment containing only the Ig κ constant-region gene and the light-chain enhancer.

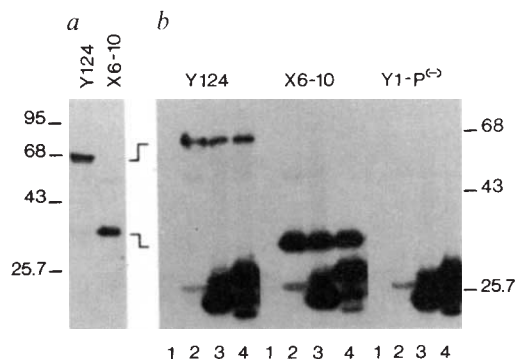


Fig. 2 Western blot characterization of the secreted chimaeric CD4 proteins. Stable transformants were obtained by transfecting CD4 constructs into X63-0 cells by protoplast fusions, with subsequent selection with mycophenolic acid and xanthine as described previously¹⁷. Y124, X6-10, Y1-p(-) are transformants obtained with pHT4-Y1, pHT4-X6 and pHT4-Y1-p(-) constructs, respectively. *a*, Cell culture supernatants and *b*, immunoprecipitations of them using either monoclonal anti-CD4 antibodies, anti-mouse κ antibodies, or purified protein A from *Staphylococcus aureus* as a negative control. All antibodies and protein A were immobilized on Sepharose 4B. 1, protein A; 2, sheep anti-mouse κ ; 3, monoclonal mouse anti-CD4 10A12; 4, monoclonal mouse anti-CD4, 6D10. The monoclonal antibody 6D10 has been selected for HIV neutralizing activity, whereas 10A12 recognizes a different, non-neutralizing CD4 epitope. The results were visualized by using radioactive anti- κ antibodies in Western blots. The strong radioactivity at the bottom of the gel is due to the light chains of the mouse monoclonal antibodies eluted from the Sepharose when samples were prepared for electrophoresis. Standard molecular weight markers are shown (K) to the left (*a*) and right (*b*).

100 ng ml⁻¹ respectively). It is interesting that no binding of the mouse CD4 analogue L3T4, when solubilized in a similar way to CD4 (A.T. *et al.*, in preparation), could be detected even at 100 times-increased concentrations, consistent with an inability to neutralize HIV *in vitro* (see below).

The CD4- κ chimaeric molecules were purified from cell cul-

Fig. 3 Soluble CD4 binding to purified HIV-1 envelope protein gp120. Purified gp120 was spotted at different concentrations on nitrocellulose strips and incubated with various culture supernatants (5 ml) overnight at 4 °C and the binding detected with radioactive anti- κ antibodies. X63, supernatant from non-transfected X63 cell line; L3T4, X6-10 and Y124 supernatants from transfectants producing soluble L3T4 (10 μ g ml⁻¹), truncated CD4 (100 ng ml⁻¹) and 'complete' CD4 (20 ng ml⁻¹) chimaeric proteins respectively. Concentrations were estimated from the κ content of the various supernatants by using κ -specific radioimmunoassay.

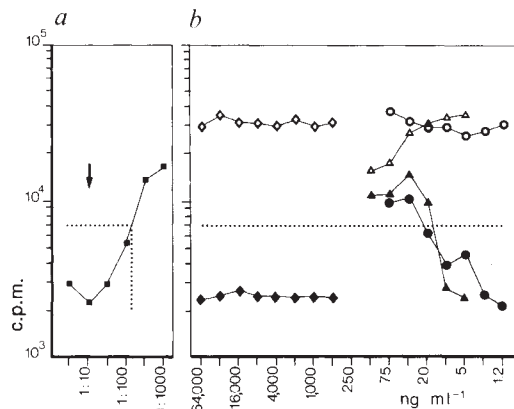


Fig. 4 Effect of solubilized CD4 on the infectivity of HIV-1. *a*, Titration of the stock virus. The arrow shows the virus dose used in the neutralization assay. The dotted line indicates 75% reduction of viral infectivity. *b*, Neutralization assay performed as described above. Open symbols represent the controls where the inhibitors are added to the assay without the virus, and filled symbols with the virus. (◆) L3T4, (▲) X6-10, (●) Y124.

Methods. The neutralization test was essentially performed as described^{19,20}. The assay uses HTLV-1 transformed MT-4 cells, which are highly sensitive to HIV-1 cytopathogenicity²¹. Reduction of viral infectivity was quantitated by measuring [³H]thymidine uptake as a marker for cell proliferation. For neutralization of the virus, the appropriate predilution of CD4 (100 μ l) was mixed with an equal volume of the virus infectious dose (1:10 dilution of the stock virus). The mixture (50 μ l) was added to an equal volume of MT-4 cell suspension (6×10^5 cells ml⁻¹) in 96 well microtitre-plates. Cells were cultured in Click-RPMI supplemented with 20% inactivated FCS and 25 mM HEPES at 37 °C in a 5% CO₂ atmosphere. After 3 d, the cells were fed with 100 μ l fresh medium. Four days after infection, 0.1 μ Ci [³H]thymidine was added to the cells. Twenty hours later, the cells were collected on glass fibre filters and washed with 10% TCA. The radioactivity of the filters was determined. All tests were done in triplicate.

ture supernatants with monoclonal rat anti-mouse κ affinity columns¹⁷ and used as neutralizers for HIV infectivity *in vitro*¹⁹⁻²¹ (Fig. 4). Both types of CD4 proteins were equally effective at preventing virus-dependent cytopathic effects, whereas the L3T4 mouse analogue did not inhibit even at 500 times-increased concentrations. Because both forms of soluble CD4 were neutralizing even at concentrations of 30 ng ml⁻¹, we conclude that the immunoglobulin-like and second domains of CD4 are sufficient to protect target cells against HIV infection. Also the lack of N-linked glycans in the two-domain molecules suggests that these sugar residues have little or no importance in CD4/HIV gp120 interaction.

We have shown here that the two N-terminal domains of CD4 in soluble form are able to inhibit HIV infection of target cells. It remains to define the regions of the CD4 molecule necessary for its normal physiological function on T lymphocytes; if the areas of CD4 which interact with HIV gp120 glycoprotein are

separate from those which bind its natural ligands, then this would present new possibilities for AIDS therapy.

We would like to thank Ms Yvonne Schneemann for excellent technical assistance, Ms Judie Hossmann for preparing the manuscript, and Drs Alan Tunnacliffe and Eddy Roosnek for critical reading. We would also like to thank Dr H. Kiefer for CD4-specific oligonucleotide preparation, Dr S. Ryser for providing human genomic libraries and Dr J. Mous for helpful discussions. Purified gp120 was a gift from Dr P. Berman and the anti-CD4 monoclonal antibodies were provided by Dr R. van Lier. The Basel Institute for Immunology was founded and is supported by F. Hoffman-La Roche, Ltd Co., Basel, Switzerland.

Received 9 November; accepted 13 November 1987.

- Curran, J. *et al. Science* **229**, 1352-1357 (1985).
- Weiss, R. *et al. Nature* **324**, 572-575 (1986).
- Dalgleish, A. *et al. Nature* **312**, 763-766 (1984).
- Klatzmann, D. & Gluckman, J. C. *Immunol. Today* **7**, 291-297 (1986).
- McDougal, J. *et al. J. Immun.* **135**, 3151-3162 (1985).
- Reinherz, E. L. & Schlossman, S. F. *Cell* **19**, 821-827 (1980).
- Klatzmann, D. *et al. Nature* **312**, 767-768 (1984).
- McDougal, J. *et al. J. Immun.* **137**, 2937-2944 (1986).
- Maddon, P. J. *et al. Cell* **47**, 333-348 (1986).
- Lasky, L. A. *et al. Cell* **50**, 975-985 (1987).
- Starich, B. R. *et al. Cell* **45**, 637-648 (1986).
- Alizon, M., Wain-Hobson, S., Montagnier, L. & Sonigo, P. *Cell* **46**, 63-74 (1986).
- Wiley, R. *et al. Proc. Natn. Acad. Sci. U.S.A.* **83**, 5038-5042 (1986).
- Hahn, B. *et al. Science* **232**, 1548-1553 (1986).
- Maddon, P. J. *et al. Cell* **42**, 93-104 (1985).
- Littman, D. R. & Gettner, S. N. *Nature* **325**, 453-455 (1987).
- Trautwein, A., Dolder, B. & Karjalainen, K. *Eur. J. Immun.* **16**, 851-854 (1986).
- Classon, J. B., Tsagaratos, J., McKenzie, I. F. & Walker, I. D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4499-4503 (1986).
- Harada, S., Purtilo, D. T., Koyanagi, Y., Sonnabend, J. & Yamamoto, N. *J. immun. Meth.* **92**, 177-181 (1986).
- Harada, S., Yoshiyama, H. & Yamamoto, N. *J. Clin. Microbiol.* **22**, 908-911 (1985).
- Harada, S., Koyanagi, Y. & Yamamoto, N. *Science* **229**, 563-566 (1985).

ICAM-1 a ligand for LFA-1-dependent adhesion of B, T and myeloid cells

Malegapuru W. Makgoba*, Martin E. Sanders*,
Gale E. Ginther Luce*, Michael L. Dustin†,†,
Timothy A. Springer†, Edward A. Clark‡,†,
Patrice Mannoni§ & Stephen Shaw*

* Immunology Branch, NCI, NIH, Bethesda, Maryland 20892, USA

† Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA

‡ Regional Primate Research Center, Washington University, Seattle, Washington 98195, USA

§ Inserm Unit 119, 27 Boulevard Lei Roure, Marseille 13009, France

Cell-cell adhesion is essential for many immunological functions¹⁻⁴. The LFA-1 molecule, a member of a superfamily of adhesion molecules, participates in adhesion which is critical to the function of each of the three major subsets of leukocytes: lymphocytes, monocytes and granulocytes⁵. Putative LFA-1 ligands have been identified functionally in different laboratories using three different monoclonal antibodies that inhibit LFA-1-mediated leukocyte adhesion in particular model systems; however, there may be more than one LFA-1 ligand⁶⁻⁸. We have directly compared the three relevant monoclonal antibodies, and show that each binds to the same molecule, intercellular-adhesion molecule-1 (ICAM-1). Most important, B, T and myeloid cells adhere specifically to purified ICAM-1-coated surfaces; such adhesion has distinctive requirements for Mg²⁺ and Ca²⁺. This constitutes biochemical evidence that ICAM-1 functions as a ligand for LFA-1-dependent adhesion by a variety of leukocytes.

Recent functional studies indicate that ICAM-1, recognized by the monoclonal antibody (mAb) RR1/1, is a ligand for LFA-1 in several models of antigen-independent B- and T-cell adhesion and in cytotoxic T-cell recognition⁶⁻⁹. Several other mAbs have

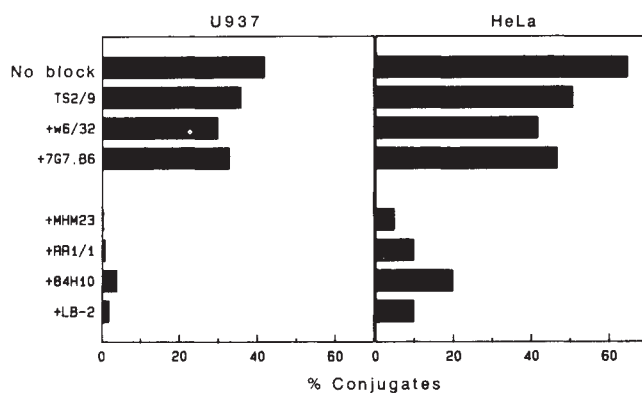


Fig. 1 Inhibition of LFA-1-dependent conjugate formation. Conjugate formation was measured between the cytotoxic T-cell clone 8.2 and two antigen-negative targets, U937 and HeLa. Conjugates were enumerated by two-colour cytometry as described^{15,23,24}. To isolate the LFA-1 pathway of adhesion the LFA-3 mAb TS2/9 was continuously present in all assays (except the 'no block') in saturating concentrations to inhibit adhesion via the CD2/LFA-3 pathway^{15,24}. Antibodies used were LFA-3 mAb TS2/9 (ref. 25), HLA class I mAb w6/32 (ref. 26), IL-2 mAb 7G7.B (ref. 27), LFA-1 β mAb MHM23 (ref. 28), RR1/1, 84H10 and LB-2. Antibodies were continuously present during the assay at 100 $\mu\text{g ml}^{-1}$ purified IgG except for MHM23 and TS2/9 which were present as 300 $\mu\text{g ml}^{-1}$ of Fab fragment.

recently been identified with properties suggesting that they also define molecules which function as ligands for LFA-1 in adhesion by different kinds of leukocytes. These include: (1) LB-2, a B-cell activation marker¹⁰ which inhibits both B- and myeloid homotypic-cell adhesion¹¹; and (2) 84H10, which was identified by screening for preferential binding to myeloid leukaemic cells and subsequently shown to inhibit the adhesion of such cells to bone-marrow stromal cells (P.M., unpublished data). Furthermore 84H10 has been shown to inhibit homotypic-B-cell aggregation¹². Here we use three different approaches which show that these three mAbs bind a single molecule (ICAM-1) and that purified ICAM-1 functions as a ligand for LFA-1 in cell adhesion.

We compared the ability of the three mAbs to inhibit LFA-1-dependent conjugate formation between a cytotoxic T-cell clone and two antigen-negative targets of different lineage: a myeloid line (U937) and a cervical-carcinoma line (HeLa) (Fig. 1). Each of the putative ICAM-1 antibodies inhibited LFA-1-dependent conjugate formation with each target cell but control mAbs did not. The residual conjugate formation between the T-cell clone and HeLa cells in the presence of each putative ICAM-1 mAbs may reflect the existence of an additional LFA-1 ligand on HeLa cells as previously proposed^{6,8}.

To determine whether LB-2, 84H10 and RR1/1 mAbs bind the same molecule, each was assayed for its ability to bind the molecule immunoaffinity-purified on an 84H10 column (Fig. 2). All three antibodies bound the purified 84H10 antigen even at concentrations of 5 ng ml⁻¹; the specificity of binding was shown by lack of binding by irrelevant LFA-3 mAb. Results of competitive-binding studies indicate that RR1/1 binds to an epitope that is spatially separate from that to which 84H10 and LB-2 bind (data not shown). Thus, RR1/1, LB-2 and 84H10 bind to epitopes on the same molecule, which we call ICAM-1 as proposed by Rothlein *et al.*⁶. This finding unifies observations from multiple laboratories regarding the function, structure, tissue distribution and genetics of ICAM-1. Most important, functional understanding of ICAM-1 is broadened by correlating the data reported for RR1/1, LB-2 and 84H10, and by our studies using purified ICAM-1 (see below). RR1/1 detects a single chain varying in its relative molecular mass (M_r) from 90 to 114K⁶, LB-2 detects a chain of 76K¹⁰ or 84K¹¹ and 84H10 to a chain of 105K¹³. This variation is consistent with results

§ Present address: Department of Chemical Pathology, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London W12 0HS, UK.

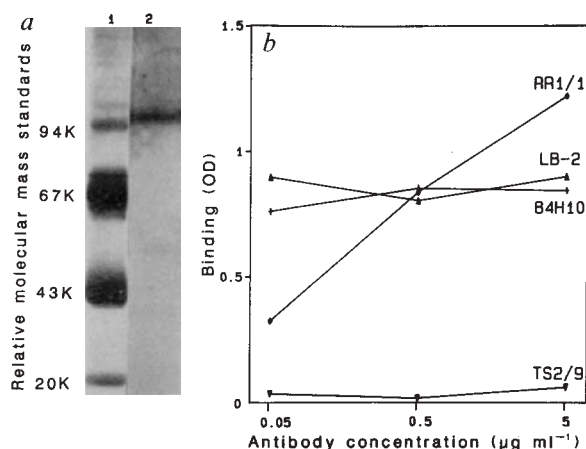


Fig. 2 Binding of three mAb to purified ICAM-1. Triton X-100 solubilized material from 3×10^8 cells of the L428 Reed-Sternberg line²⁹ was purified on a column of 84H10 mAb coupled to Sepharose-4B beads. After washing the column, material was eluted at pH 2.5 in 0.01 M Tris/0.05 M glycine buffer. The eluate ran as a single band of 95–97K on a 12.5% SDS-polyacrylamide gel electrophoresis (PAGE) (a, lane 2). a, Lane 1 shows relative molecular-mass standards. Purified ICAM-1 was applied to glutaraldehyde/poly-L-lysine-treated microtitre plates as follows: 96-well Nunc-Immunoplates were treated with 100 μ l per well of a 0.2% glutaraldehyde in 0.1 M sodium carbonate-HCl pH 9 for 1 h at room temperature. The plate was washed twice with distilled water. 50 μ l of poly-L-lysine (50 μ g per ml in 0.05 M sodium bicarbonate) were added to each well and left to incubate for 2 h. The plate was washed twice with 0.05 M sodium bicarbonate. To each well 25 μ l of purified ICAM-1 and 25 μ l of 0.2% glutaraldehyde (in 0.05 M sodium bicarbonate) were added. Incubation was allowed to proceed for 8 h. Unbound sites were neutralized by filling each well with 0.05 M Tris-HCl pH 7.8 plus 2.5% human serum albumin (HSA) for 2 h. The plate was washed twice with Tris-HSA. Antibody binding to the immobilized antigen b) was measured by enzyme-linked immunosorbent assay (ELISA) using purified mAb (IgG) as the first step, β -galactosidase-conjugated F(ab')₂ sheep anti-mouse IgG (heavy plus light chains) as the second step, and measuring the rate of conversion of *p*-nitrophenyl- β -galactoside to a coloured product (using Bethesda Research Laboratories hybridoma screening reagent). Data are given as the average of the optical density at 405 nm of duplicate wells after subtraction of the OD in the absence of a first-step antibody; standard errors were less than 10% of the mean. No specific binding was observed in control wells, in which antibodies were incubated in HSA-coated wells (data not shown).

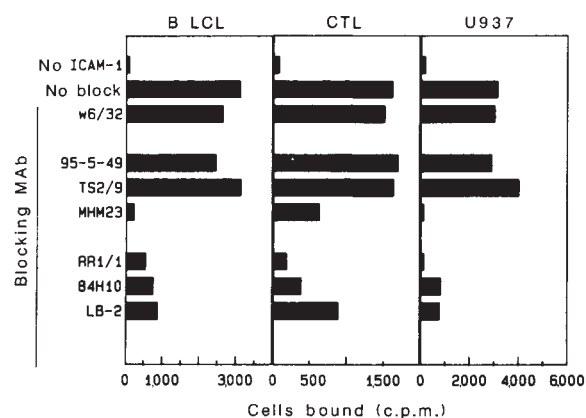


Fig. 3 Antibody inhibition of cell binding to purified ICAM-1 immobilized on microtitre plates. Plates were prepared as described in the legend to Fig. 2. Control wells (designated 'no ICAM-1') were coated with HSA; all other wells were coated with purified ICAM-1. Cells of three different leukocyte lineages were labelled with ⁵¹Cr and 10^5 cells added to each well in 100 μ l of phosphate-buffered saline (PBS) 0.5% HSA, and incubated for 20 min at 37 °C. Antibodies being tested for their ability to inhibit cell binding were present continuously at 50 μ g ml⁻¹ final concentration. The plates were washed gently three times with PBS/0.5% HSA to remove unbound cells. The bound cells were lysed with 5% Triton X-100 and the lysate counted for radioactivity. Results are expressed as the mean c.p.m. of duplicate wells; standard errors are all less than 15% of the mean. Antibodies used were as described in Fig. 1 legend, and CD2 mAb 95-5-49 (ref. 30).

showing that ICAM-1 is synthesized from a precursor of 73K, of which 55K is protein, and that cell-type-dependent differences in carbohydrate processing account for variation in *M_r*. Staining shows that ICAM-1 is present in a variety of cells in tissue sections, particularly endothelium^{6,9}, a pattern consistent with the staining pattern of 84H10 (ref. 13). Our analysis of staining on a variety of *in vitro* cell lines shows complete concordance in binding of the antibodies RR1/1, LB-2 and 84H10 (M.W.M., in preparation). Enhancement of antibody binding following cell activation is characteristic not only of LB-2 (ref. 11), but also of RR1/1 (ref. 9) and 84H10 (data not shown). The gene coding for ICAM-1 maps to chromosome 19, based on previous studies with the LB-2 mAb¹⁰.

Inhibition of LFA-1-dependent binding by ICAM-1 mAb RR1/1 at the target level^{8,9} has been the basis for the inference that ICAM-1 is a ligand for LFA-1. But alternative explanations are possible, for example that binding of ICAM-1 mAb delivers an inhibitory signal which prevents target participation in LFA-1-dependent adhesion. To test directly the hypothesis that ICAM-1 is a ligand involved in LFA-1-dependent adhesion, we studied cell adhesion to purified ICAM-1 (Fig. 3). LFA-1-positive cells of three different lineages (a lymphoblastoid B cell, a cytotoxic T-cell clone and a promonocytoid cell) bound to

ICAM-1 immobilized on plastic. Three lines of evidence show that this adhesion is specific: (1) the adhesion is inhibited by the three ICAM-1 mAbs and the LFA-1 mAbs but not by the LFA-3, CD2 and class I histocompatibility locus antigen (HLA) mAbs; (2) the cells did not adhere to other proteins immobilized on plastic (for instance albumin, Fig. 3); (3) cells like L428 that lack LFA-1 expression did not bind to plastic-immobilized ICAM-1 (data not shown). These results confirm and extend studies using purified ICAM-1 from a RR1/1 affinity column, incorporated into phospholipid membranes, (S.D.M. and T.A.S. Cell, in press). These studies show directly that ICAM-1 is a ligand for LFA-1-dependent cell adhesion. The simplest interpretation of these data is that ICAM-1 is a ligand for LFA-1. However, the possibility that ICAM-1 is a ligand for some other adhesion receptor whose function depends on LFA-1 cannot be excluded.

LFA-1-dependent adhesion has been shown previously to require divalent cations^{14,15}. We analysed the cation requirements of LFA-1-dependent adhesion to purified ICAM-1 and directly compared them with the cation requirements of intercellular adhesion via the molecules LFA-1 and ICAM-1 (Fig. 4). The similarities between the results of the two assay systems are striking: Mg²⁺ alone is sufficient for optimal adhesion; Ca²⁺ alone is not, but can synergize with suboptimal concentrations of Mg²⁺. For example, 1 mM Ca²⁺ supports minimal (<10% of optimal) adhesion in each system, however, in the presence of 0.1 mM Mg²⁺, 1 mM Ca²⁺ enhances adhesion from ~30% to ~90% of the maximal adhesion. The similarities between the two assay systems emphasize that the present assay of cell binding to purified ICAM-1 faithfully reflects the biology of LFA-1 interactions which are involved in more complex intercellular interactions. This pattern of cation requirement is characteristic of LFA-1-dependent adhesion and not observed, for example, with adhesion of the same cells via CD2/LFA-3 interaction¹⁵. These data indicate that divalent cations are required for the interaction of LFA-1 with ICAM-1. Two divalent-cation-binding sites may be involved: one requiring Mg²⁺, which is saturated at Mg²⁺ concentrations below 0.1 mM, and another which binds either Mg²⁺ or Ca²⁺ but with somewhat lower affinity than the first site.

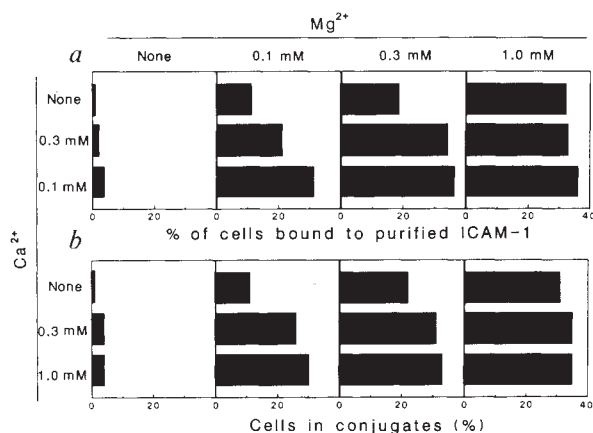


Fig. 4 Distinctive divalent-cation requirements of LFA-1-dependent adhesion to ICAM-1. The ability of Ca^{2+} and Mg^{2+} to support cell binding to purified ICAM-1 (a) was compared with the requirement for Ca^{2+} and Mg^{2+} in intercellular adhesion (b). Binding to ICAM-1 was measured as outlined in Fig. 3 and T-cell conjugates as described in Fig. 1 except that the buffer used in both assays was HEPES-buffered saline consisting of 0.13 M NaCl, 10 mM HEPES, 2 mg ml⁻¹ glucose, 5 mg ml⁻¹ HSA. MgCl_2 and CaCl_2 were added at the concentrations indicated during the assay. Cells used were B LCL in panel a and T-cell clone 8.2 with target U937 in panel b; U937 was chosen as the target cell because it uses ICAM-1 almost exclusively as a ligand in LFA-1-dependent intercellular adhesion (see also Fig. 1).

Multiple lines of evidence show that ICAM-1 is a ligand for LFA-1-dependent adhesion of multiple lineages of haematopoietic cells: T cells, B cells and myeloid cells bind to purified ICAM-1 (Fig. 3); RRI/1 inhibits binding of T, B and myeloid cells^{6,9}; LB-2 mAb inhibits adhesion of B and myeloid cells¹¹; 84H10 inhibits binding of myeloid cells (P.M., unpublished) and B cells¹². In normal tissue sections, staining is most intense on endothelium but is increased on many cells by such factors as inflammation, exposure to cytokines and neoplastic transformation⁵. Thus, the role of ICAM-1 as an LFA-1 ligand may be critical in immune surveillance during inflammation or neoplastic transformation.

Structural information on LFA-1 shows that it is part of a large family of cell-interaction molecules highly conserved throughout evolution. LFA-1 and the other molecules with which it shares a β chain (MAC-1 and p150/95) are restricted in their distribution to haematopoietic cells⁵ but they are closely related to more widely-distributed families of mammalian adhesion molecules and even to the *Drosophila* position-specific protein which is critical for cell positioning during embryogenesis and morphogenesis¹⁶⁻²². It remains to be shown whether ICAM-1, like other ligands for many of these adhesion receptors, has a critical RGD peptide in its sequence and whether it can be used as a ligand by more than one receptor.

T.A.S. is supported by the NIH. We thank Drs R. E. Gress, J. E. K. Hildreth, W. E. Biddison and R. Quinones for mAbs and Drs W. E. Biddison and P. Henkart for useful discussion.

Received 11 August; accepted 6 November 1987.

- Rouse, R. V., Reichert, R. A., Gallatin, W. M., Weissman, I. L. & Butcher, E. C. *Am. J. Anat.* **170**, 391-405 (1984).
- Bjerknes, M., Cheng, H. & Ottaway, C. A. *Science* **231**, 402-405 (1986).
- Bell, G. I. *Immun. Today* **4**, 237-240 (1983).
- Inaba, K. & Steinman, R. M. *J. exp. Med.* **163**, 247-261 (1986).
- Springer, T. A., Dustin, M. L., Kishimoto, T. K. & Marlin, S. D. *A. Rev. Immun.* **5**, 223-252 (1987).
- Rothlein, R., Dustin, M. L., Marlin, S. D. & Springer, T. A. *J. Immun.* **137**, 1270-1274 (1986).
- Makgoba, M. W. *et al. Ann. N.Y. Acad. Sci.* (in the press).
- Makgoba, M. W. *et al. Eur. J. Immun.* (in the press).
- Dustin, M. L., Rothlein, R., Bhan, A. K., Dinarello, C. A. & Springer, T. A. *J. Immun.* **137**, 245-254 (1986).
- Clark, E. A., Ledbetter, J. A., Holly, R. C., Dinndorf, P. A. & Shu, G. *Hum. Immun.* **16**, 100-113 (1986).

- Patarroyo, M., Clark, E. A., Prieto, J., Kantor, C. & Gahmberg, C. G. *FEBS Lett.* **210**, 127-131 (1987).
- Marlin, S. D., Miller, L. M. & Springer, T. A. in *Leucocyte typing III. White cell differentiation antigens* (eds McMichael, A. J. *et al.*) 832-833 (Oxford University Press, New York, 1987).
- Hogg, N. & Horton, M. A. in *Leucocyte Typing III. White Cell Differentiation Antigens* (eds McMichael, A. J. *et al.*) 576-602 (Oxford University Press, Oxford, 1987).
- Martz, E. *J. Cell Biol.* **84**, 584-598 (1980).
- Shaw, S. *et al. Nature* **323**, 262-264 (1986).
- Hynes, R. O. *Cell* **48**, 549-554 (1987).
- Lawler, J. & Hynes, R. O. *J. Cell Biol.* **103**, 1635-1648 (1986).
- Suzuki, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 8614-8618 (1986).
- Takada, Y., Strominger, J. L. & Hemler, M. E. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3239-3243 (1987).
- Law, S. K. A. *et al. EMBO J.* **6**, 915-919 (1987).
- Miller, L. J., Wiebe, M. & Springer, T. A. *J. Immun.* **138**, 2381-2383 (1987).
- Kishimoto, T. K., O'Connor, K., Lee, A., Roberts, T. M. & Springer, T. A. *Cell* **48**, 681-690 (1987).
- Luce, G. G., Gallop, P. M., Sharrow, S. O. & Shaw, S. *BioTechniques* **3**, 270-272 (1985).
- Shaw, S. & Luce, G. G. *J. Immun.* **139**, 1037-1045 (1987).
- Sanchez-Madrid, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 7489-7493 (1982).
- Barnstable, C. J. *et al. Cell* **14**, 9-20 (1978).
- Rubin, L. A., Kurman, C. C., Biddison, W. E., Goldman, N. D. & Nelson, D. L. *Hybridoma* **4**, 91-102 (1985).
- Hildreth, J. E. & August, J. T. *J. Immun.* **134**, 3272-3280 (1985).
- Schaadt, M., Diehl, V., Stein, H., Fonatsch, C. & Kirchner, H. H. *Int. J. Cancer* **26**, 723-731 (1980).
- Quinones, R. R., Lunney, J. & Gress, R. E. *Transplantation* (in the press).

Aggregation of chromaffin granules by calpactin at micromolar levels of calcium

Debra S. Drust & Carl E. Creutz

Department of Pharmacology, University of Virginia, Charlottesville, Virginia 22908, USA

Several cytosolic proteins bind to secretory granule membranes in a Ca^{2+} -dependent manner and thus may be involved in the mediation of membrane interactions during exocytosis^{1,2}. One of these proteins³, calpactin, is a tetramer consisting of two heavy chains of relative molecular mass (M_r) 36K (p36) and two light chains of 10K (p10)^{4,5}. We report here that calpactin promotes the Ca^{2+} -dependent aggregation and fatty acid-dependent fusion of chromaffin granule membranes at a level of Ca^{2+} that is lower than that reported for other granule-aggregating proteins, and which parallels the Ca^{2+} requirement for secretion from permeabilized chromaffin cells^{6,7}. We found subunits of calpactin to be inactive in promoting granule aggregation. Two distinct 33K proteolytic fragments of p36, differing at their N termini, also promote granule aggregation but with different Ca^{2+} sensitivities from calpactin. These differences suggest that the N-terminal portion of p36 modulates the Ca^{2+} /lipid binding sites in the core portion of p36 (ref. 5).

The heavy chain of calpactin, p36, was originally isolated as a major substrate for tyrosine kinases *in vivo* and *in vitro*⁸⁻¹⁰, and it can also be phosphorylated by protein kinase C^{11,12}. The main sites of phosphorylation for these kinases are at the N terminus of p36 (Tyr 23 and Ser 25)^{11,13}. Little was known about the possible function of p36 until it was discovered that p36 is a subunit of a prominent Ca^{2+} -dependent membrane-binding protein (protein I) present in intestinal epithelium⁴. In addition to binding to membranes, the p36₂-p10₂ complex binds to actin in the presence of Ca^{2+} (ref. 4) and was therefore called calpactin I¹⁴. The sequence of p36 has recently been found to be ~50% identical to lipocortin, a 37K phospholipase A₂ inhibitor isolated from placenta¹⁵⁻¹⁸.

A group of 23 cytosolic proteins, called chromobindins¹, have been isolated based on their ability to bind to secretory vesicle (chromaffin granule) membranes of the adrenal medulla in a Ca^{2+} -dependent way^{1,2}, and represent candidates for mediators of membrane interactions in exocytosis. By peptide mapping and immunological cross-reactivity, chromobindin No. 8 was identified as p36 (ref. 3). This is one of a subgroup of chromobin-

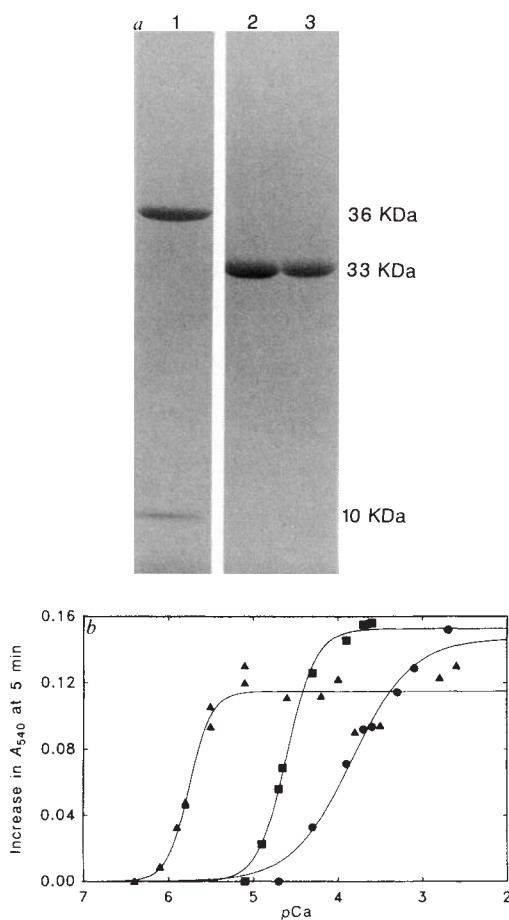
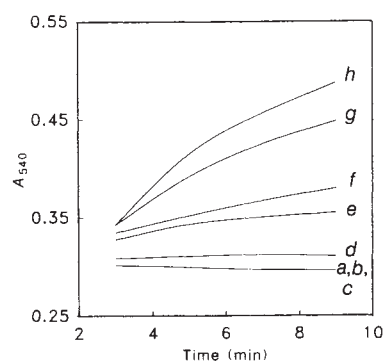


Fig. 1 Ca^{2+} -dependent aggregation of bovine chromaffin granules promoted by calpactin and proteolytic fragments of the heavy chain of calpactin, p33A and p33B. *a*, SDS-polyacrylamide gels of calpactin (lane 1), p33A (lane 2) and p33B (lane 3) isolated as described below. *b*, Ca^{2+} -titration curves for granule aggregation by calpactin ($\blacktriangle$), p33B ($\blacksquare$) and p33A ($\bullet$).

Methods. Calpactin was purified from bovine lung according to Glenney²² with the following modifications. Buffers contained 0.25 μM phenylmethylsulphonyl fluoride (PMSF), 1 $\mu\text{g ml}^{-1}$ aprotinin, and 25 mM HEPES (pH 7.4). Lung was homogenized in 25 mM HEPES (pH 7.4), 300 mM sucrose, 2.5 mM EGTA and centrifuged for 15 min at 2,900g. 5 mM CaCl_2 was added to the supernatant which was allowed to stand for 10 min at 4°C, and then centrifuged at 27,000g for 45 min. Gel filtration was on Ultrogel ACA 44 rather than Sephacryl S200. Fractions containing calpactin were dialysed against 25 mM HEPES (pH 7.4), 1 mM NaN_3 , 0.5 mM DTT and chromatographed on a DEAE-sephacel column. The flow-through was adjusted to pH 6.0 and applied to an FPLC Mono S column (Pharmacia) equilibrated in 25 mM 2[*N*-morpholino]ethanesulphonic acid (MES) (pH 6.0), 300 mM sucrose. Calpactin eluted at about 0.3 M KCl. p33A and p33B were purified from bovine intestinal epithelium following ref. 4 with the following modifications. Intestinal mucosa was scraped off into 50 mM sucrose, 2.5 mM EGTA, 10 mM HEPES (pH 7.4), 0.25 mM PMSF and then homogenized in a Waring blender for 2 min. After centrifugation at 2,900g for 15 min, 10 mM CaCl_2 was added to the supernatant which was allowed to stand for 10 min at 4°C, and then centrifuged at 27,000g for 45 min. The pellet was resuspended and washed as described⁴ using HEPES instead of imidazole. The final pellet was resuspended in 5 mM EGTA, 0.5 mM DTT, 1 mM NaN_3 , 25 mM HEPES (pH 7.4), centrifuged at 100,000g for 1 h, and the supernatant applied to a DEAE-sephacel column and an FPLC Mono S column as described above for calpactin. p33A eluted at 0.22 M KCl and p33B eluted at 0.27 M KCl. Granule aggregation was monitored as an increase in turbidity (A_{540}) in 5 min as described¹⁹ at room temperature in a buffer containing 30 mM KCl, 240 mM sucrose, 40 mM HEPES (pH 7.0). Free Ca^{2+} levels were maintained with Ca^{2+} /EGTA buffers¹⁹ and measured using a Ca^{2+} electrode ($p\text{Ca} = -\log [\text{Ca}^{2+}]$). Protein concentration²⁷ was 12.5 $\mu\text{g ml}^{-1}$, 10 $\mu\text{g ml}^{-1}$, 7 $\mu\text{g ml}^{-1}$ for calpactin, p33A and p33B respectively. Concentrations were determined to give maximal levels of aggregation. Aggregation data were fitted to the equation $y = mkx^n / (1 + kx^n)$, (m , maximum; k , constant; n , Hill coefficient).

Fig. 2 Time course of chromaffin granule aggregation by calpactin subunits and the reconstituted tetramer. Dissociated subunits of calpactin, p36 and p10, were tested for the ability to promote granule aggregation at 3 μM Ca^{2+} and 1 mM Ca^{2+} (*a*, p10 at 3 μM Ca^{2+} ; *b*, p10 at 1 mM Ca^{2+} ; *c*, p36 at 3 μM Ca^{2+} ; *d*, p36 at 1 mM Ca^{2+}). The same amounts of p36 (16 μg) and p10 (4.4 μg) were incubated together at room temperature for 30 min and added to granules. Granule aggregation was measured at 3 μM Ca^{2+} (*e*) and 1 mM Ca^{2+} (*h*). For comparison, granule aggregation at 12 μg of native calpactin at 3 μM Ca^{2+} (*f*) and 1 mM Ca^{2+} (*g*) are also shown. Aggregation in the absence of added proteins at 3 μM or 1 mM Ca^{2+} is also represented by the bottom trace (*a*, *b*, *c*).

Methods. Calpactin isolated as described in Fig. 1 legend was concentrated and dissociated into subunits by addition of concentrated stocks of guanidine-HCl (pH 5.5) and DTT to final concentrations of 6 M guanidine-HCl and 1 mM DTT. The separated subunits were chromatographed on an FPLC Superose-12 column in a buffer containing 6 M guanidine-HCl and 1 mM DTT. The fractions containing p36 and p10 were dialysed against 1 mM DTT, 30 mM KCl, 240 mM sucrose, 40 mM HEPES (pH 7.0).



dins that will bind to lipids extracted from the granule membrane or to pure phosphatidylserine vesicles¹. As the heavy chain of calpactin can associate with chromaffin granule membranes in a Ca^{2+} -dependent way, we sought to determine if calpactin can also promote granule aggregation and fusion. Not all granule membrane-binding proteins exhibit such a 'bivalent' activity, which is characteristic of the protein synexin¹⁹ and has been reported for calelectrin^{3,20}. To isolate large quantities of calpactin we first tried to purify this protein from bovine intestinal epithelium using a modified procedure of Gerke and Weber⁴ (see legend to Fig. 1). The p36 subunit is susceptible to proteolysis during isolation, resulting in a 3K N-terminal 'tail' and a

33K 'core' to which p10 cannot bind^{5,21}. Because of this problem we obtained primarily two proteolytic fragments of p36, p33A and p33B, which differ at the N terminus. By comparing the N-terminal sequence of these two proteins with the sequence of bovine p36 (ref. 16), we determined that p33A begins at residue 28 and p33B at residue 44. The more abundant of the two proteins, p33A, begins at or near the site of the core fragments of p36 which have been reported previously^{11,13}. Both p33A and p33B aggregate chromaffin granules, but with different Ca^{2+} sensitivities (Fig. 1). The shorter fragment, p33B, is more sensitive to Ca^{2+} , requiring 24 μM Ca^{2+} for half-maximal aggregation (K_d), whereas p33A requires 141 μM Ca^{2+} . This

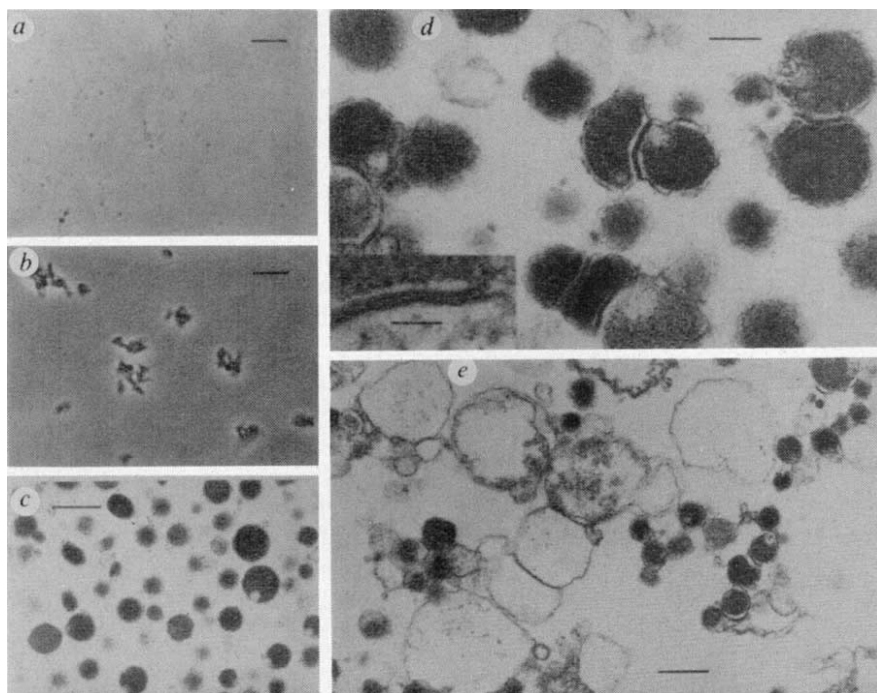


Fig. 3 Light and electron micrographs of calpactin-induced granule aggregation and fusion. Light micrographs of granules incubated without (a) and with (b) calpactin at $3 \mu\text{M}$ Ca^{2+} , pH 7.0 for 10 min. For electron micrographs (c–e) granules were allowed to aggregate for 10 min at $3 \mu\text{M}$ Ca^{2+} in the presence or absence of calpactin and then for (e) arachidonic acid ($5 \mu\text{g}$ per ml) was added for 10 min to allow fusion to occur. Granules were fixed and prepared for electron microscopy as described previously²⁴. Granules incubated in the absence (c) or presence (d) of calpactin at $3 \mu\text{M}$ Ca^{2+} , pH 7.0. Inset of d is at a higher magnification to show the junction between granules. Fusion promoted by calpactin at $3 \mu\text{M}$ Ca^{2+} , pH 6.0 is evident in (e) as large vesicles with diluted core material. Bars: a and b, $10 \mu\text{m}$; c and e, $0.5 \mu\text{m}$; d, $0.2 \mu\text{m}$; inset of d, $0.05 \mu\text{m}$.

difference in Ca^{2+} sensitivity suggests that modifications in the N-terminal region of p36 could modulate the Ca^{2+} /lipid-binding sites in the 33K core of p36.

To test further the possibility that the N terminus of p36 could modulate the Ca^{2+} sensitivity of p36, we isolated full-length p36 in its monomeric form and as a complex with p10 and determined their Ca^{2+} dependencies for promoting granule aggregation. Calpactin was isolated from bovine lung using a modification of the procedure of Glenney²² (see Fig. 1 legend). The calpactin tetramer is more sensitive to Ca^{2+} than either p33A or p33B in promoting granule aggregation, with half-maximal aggregation occurring at $1.8 \mu\text{M}$ Ca^{2+} (Fig. 1). Lipocortin purified from bovine lung (see legend to Table 1) also aggregates chromaffin granules, but with an apparent K_d of $213 \mu\text{M}$ Ca^{2+} . This higher level of Ca^{2+} is comparable to that required by other previously described proteins that exhibit a synexin-like activity (Table 1). Recently, calpactin and lipocortin have also been shown to aggregate phosphatidylserine liposomes²³.

To obtain pure preparations of p36 we dissociated calpactin into its subunits using 6 M guanidine-HCl and then isolated the subunits by gel filtration (see Fig. 2 legend). After dialysis to remove the guanidine, p36 and p10 were tested for their ability to aggregate chromaffin granules. The p10 subunit is ineffective in promoting granule aggregation even at 1 mM Ca^{2+} (Fig. 2). Aggregation was not promoted by p36 at micromolar levels of Ca^{2+} and only a small amount occurred at 1 mM Ca^{2+} . This inability of p36 to aggregate chromaffin granules is unlikely to be caused by denaturation during the isolation procedure since recombining the isolated p36 and p10 subunits restores the Ca^{2+} sensitivity of native calpactin in promoting granule aggregation (Fig. 2). This restoration of activity, however, is critically dependent on the presence of dithiothreitol (DTT), implicating a role for a free sulphhydryl group in the reactivation process. The inability of p36 (in contrast to calpactin) to aggregate membranes at micromolar levels of Ca^{2+} suggests that subunit association is critical in this process. The site of association of p10 with p36 is at the N-terminal 29 amino acids of p36 (ref. 21). This is the same portion of p36 that is missing from the proteolytic fragments of p36, p33A and p33B, which can promote granule aggregation, but with different Ca^{2+} sensitivities than calpactin. These data again suggest that

the N-terminal 3K tail of p36 is a regulatory region that can modulate the Ca^{2+} sensitivity of the lipid-binding sites contained in the core of p36.

In the studies described above we used increase in turbidity as a measure of aggregation. To show that calpactin was actually promoting membrane interactions at micromolar levels of Ca^{2+} and to determine conditions under which it promotes fusion of the granules, we examined calpactin-induced granule aggregation by light and electron microscopy. By comparing the light micrographs in Fig. 3a, b it is evident that calpactin does promote the formation of large aggregates of granules at micromolar concentrations of Ca^{2+} . To study the junctions between these granules and to determine whether calpactin can promote fusion of the granule membranes, granule aggregates were fixed and examined in the electron microscope. At $3 \mu\text{M}$ Ca^{2+} (pH 7.0) granules are aggregated by calpactin (Fig. 3d) but not fused, as gaps between membranes are clearly visible at high magnification (Fig. 3d, inset). *cis*-unsaturated fatty acids are required to promote fusion between granules that have been aggregated by synexin²⁴. Addition of arachidonic acid to granules aggregated by calpactin at $3 \mu\text{M}$ Ca^{2+} (pH 7.0) does not result in obvious fusion of granule membranes. But after the surface

Table 1 Calcium dependence of 'synexin-like' activity of several membrane-binding proteins

Protein	Threshold [Ca^{2+}]* for granule aggregation (μM)	[Ca^{2+}] for half-maximal aggregation (μM)	Ref.
Synexin	80	200	19
Calelectrin (32K)	63	282	3
Lipocortin†	45	213	‡
p33A	12.5	141	‡
p33B	7.8	24	‡
Calpactin	0.7	1.8	‡

* [Ca^{2+}] that gives 5% of maximal turbidity increase.

† Lipocortin was isolated from bovine lung using the method described for purification of p33A and p33B. Purified lipocortin eluted at about 0.1 M KCl from an FPLC Mono S column.

‡ Data from this paper.

charge on the granules has been reduced by lowering the pH to 6.0, calpactin promotes fusion at micromolar levels of Ca^{2+} in the presence of arachidonic acid (Fig. 3e). In preliminary experiments to find physiological agents that would lower the Ca^{2+} concentration necessary for fusion we found that the addition of 2 mM Mg^{2+} , which also reduces the surface charge on granules, allows fusion to occur at lower levels of Ca^{2+} ($\leq 25 \mu\text{M}$) at pH 7.0. The threshold concentration of Mg^{2+} needed to promote fusion was between 1 and 2 mM under these conditions. These results indicate that calpactin, like synexin²⁴, is not a fusogenic protein. Instead, these proteins seem to promote the formation of contacts between membranes that will undergo fusion only when the membrane structure is also perturbed by *cis*-unsaturated fatty acids (presumably made available in the cell through the activation of phospholipases).

Recent phosphorylation studies *in vitro* show that phosphorylation in the N-terminal region of p36 by serine kinases can prevent the association of p36 with p10, although it has not been determined which residues are involved²⁵. This could be a way of regulating the function of p36 since p36 complexed to p10 can promote granule aggregation at micromolar concentrations of Ca^{2+} whereas free p36 cannot. ³²P is incorporated into p36 when chromaffin cells are preincubated with ³²P-labelled inorganic phosphate and stimulated to secrete by nicotine³, but it is not known whether a net change in phosphate content of p36 occurs during this process. If the sites of altered phosphorylation were in the N terminus, this could regulate the formation of the p36-p10 complex and, as a consequence, the ability of p36 to promote membrane interactions. Alternatively, phosphorylation of free p36 could alter the ability of p36 to promote granule aggregation.

In the present model system we have examined only the fusion of chromaffin granules, a process that occurs in the cell during compound exocytosis. Calpactin, however, is in the cortical region of the chromaffin cell, near the plasma membrane²⁶. This localization, as well as the high sensitivity of calpactin to Ca^{2+} , makes this protein a strong candidate for a critical receptor of Ca^{2+} in exocytosis that promotes contact between the secretory vesicle and the plasma membrane.

We thank Bonnie Shephard for assistance with the electron microscopy studies. This study was supported by research grants from the National Institutes of Health, the Council for Tobacco Research, and by an Established Investigator Award (to C.E.C.) from the American Heart Association.

Received 3 August; accepted 5 November 1987.

- Creutz, C. E. *et al.* *J. biol. Chem.* **258**, 14664-14674 (1983).
- Geisow, M. J. & Burgoyne, R. D. *J. Neurochem.* **38**, 1735-1741 (1982).
- Creutz, C. E. *et al.* *J. biol. Chem.* **262**, 1860-1868 (1987).
- Gerke, V. & Weber, K. *EMBO J.* **3**, 227-233 (1984).
- Glenney, J. *J. biol. Chem.* **261**, 7247-7252 (1986).
- Knight, D. E. & Baker, P. F. *J. Membrane Biol.* **68**, 107-140 (1982).
- Dunn, L. A. & Holz, R. W. *J. biol. Chem.* **258**, 4989-4993 (1983).
- Radke, K., Gilmore, T. & Martin, G. S. *Cell* **21**, 821-828 (1980).
- Erickson, E. & Erickson, R. L. *Cell* **21**, 829-836 (1980).
- Cooper, J. A. & Hunter, T. *Molec. cell. Biol.* **1**, 165-178 (1981).
- Gould, K. L., Woodgett, J. R., Isacke, C. M. & Hunter, T. *Molec. cell. Biol.* **6**, 2738-2744 (1986).
- Khanna, N. C., Tokuda, M., Chong, S. M. & Waisman, D. M. *Biochem. biophys. Res. Commun.* **137**, 397-403 (1986).
- Glenney, J. R. & Tack, B. F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7884-7888 (1985).
- Glenney, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4258-4262 (1986).
- Saris, C. J., Tack, B. F., Kristensen, T., Glenney, J. R. & Hunter, T. *Cell* **46**, 202-212 (1986).
- Kristensen, T. *et al. Biochemistry* **25**, 4497-4503 (1986).
- Huang, K.-S. *et al. Cell* **46**, 191-199 (1986).
- Wallner, B. P. *et al. Nature* **320**, 77-81 (1986).
- Creutz, C. E., Pazoles, C. J. & Pollard, H. B. *J. biol. Chem.* **253**, 2858-2866 (1978).
- Sudhof, T. C., Ebbecke, M., Walker, J. H., Fritsche, U. & Boustead, C. *Biochemistry* **23**, 1103-1109 (1984).
- Glenney, J. R., Boudreau, M., Galyean, R., Hunter, T. & Tack, B. *J. biol. Chem.* **261**, 10485-10488 (1986).
- Glenney, J. R. *Expl. Cell Res.* **162**, 183-190 (1986).
- Glenney, J. R., Tack, B. & Powell, M. A. *J. Cell Biol.* **104**, 503-511 (1987).
- Creutz, C. E. *J. Cell Biol.* **91**, 247-256 (1981).
- Johnsson, N., Van, P. N., Soling, H.-D. & Weber, K. *EMBO J.* **5**, 3455-3460 (1986).
- Burgoyne, R. D. & Cheek, T. R. *Biosci. Rep.* (in the press).
- Bradford, M. M. *Analyt. Biochem.* **72**, 248-254 (1976).

Identification of a new class of steroid hormone receptors

Vincent Giguère*, Na Yang*†, Prudimar Segui* & Ronald M. Evans*

* Howard Hughes Medical Institute, Gene Expression Laboratory, The Salk Institute for Biological Studies, PO Box 85800, San Diego, California 92138, USA

† Department of Chemistry, University of California, San Diego, California, USA

The gonads and adrenal glands produce steroids classified into five major groups which include the oestrogens, progestins, androgens, glucocorticoids and mineralocorticoids. Gonadal steroids control the differentiation and growth of the reproductive system, induce and maintain sexual characteristics and modulate reproductive behaviour. Adrenal steroids also influence differentiation as well as being metabolic regulators. The effects of each steroid depend primarily on its specific receptors, the nature of which could therefore provide a basis for classification of steroid hormone action. The successful cloning, sequencing and expression of the human glucocorticoid (hGR) (ref. 1), oestrogen^{2,3} (hER), progesterone⁴ (hPR), and mineralocorticoid⁵ (hMR) receptors, complementary DNA, plus homologues from various species⁶⁻¹¹, provides the first opportunity to study receptor structure and its influence on gene expression. Sequence comparison and mutational analysis^{12,13} show structural features common to all groups of steroid hormone receptors. The receptors share a highly conserved cysteine-rich region which functions as the DNA-binding domain^{14,15}. This common segment allows the genome to be scanned for related gene products: hMR cDNA for example, was isolated using an hGR hybridization probe⁵. In this study, using the DNA-binding domain of the human oestrogen receptor cDNA as a hybridization probe, we have isolated two cDNA clones encoding polypeptides with structural features suggestive of cryptic steroid hormone receptors which could participate in a new hormone response system.

One approach to search for unrecognized hormone response systems is to systematically employ reduced stringency hybridization to screen recombinant DNA libraries for novel hormone receptors. The DNA-binding domain of the oestrogen receptor was used as hybridization probe in analysis of a λ gt10 human testis cDNA library, and gave three positives at a frequency of one clone per 3×10^5 recombinant phages. Nucleotide sequence analysis showed that two of these clones encode the oestrogen receptor and that the third one, spanning 2.0 kilobases (kb) and termed λ hT16, gave only partial sequence similarity. The λ hT16 clone was then used to screen human fetal kidney and adult heart cDNA libraries, to give three new clones: λ hKE4 and λ hKA1 from the kidney library both represent the same gene product as λ hT16; here referred to as hERR1 (Fig. 1a). The third clone from the cardiac cDNA library (λ hH3) is only partially related and is designated hERR2 (Fig. 1c). The sequence of hERR1 is shown in Fig. 1b. The cDNA insert from λ hKA1 contains nucleotides 179-2,430, λ hKE4 represents a rare messenger RNA splicing error, with deletion of exon 2 and insertion of intron sequences. These λ hKE4 exon/intron boundaries were confirmed by cloning and partially sequencing the encompassing genomic fragments (data not shown). The sequence surrounding the first ATG agrees with the consensus¹⁶ for a translation initiation site. An open-reading frame of 521 amino acids, predicting a polypeptide of relative molecular mass (M_r) 57,300 (57.3K), is flanked by a 775-nucleotide 3'-untranslated region.

Figure 1 also shows the nucleotide and amino-acid sequence of hERR2. The translation initiation site was assigned to the methionine codon, corresponding to nucleotides 100-102,

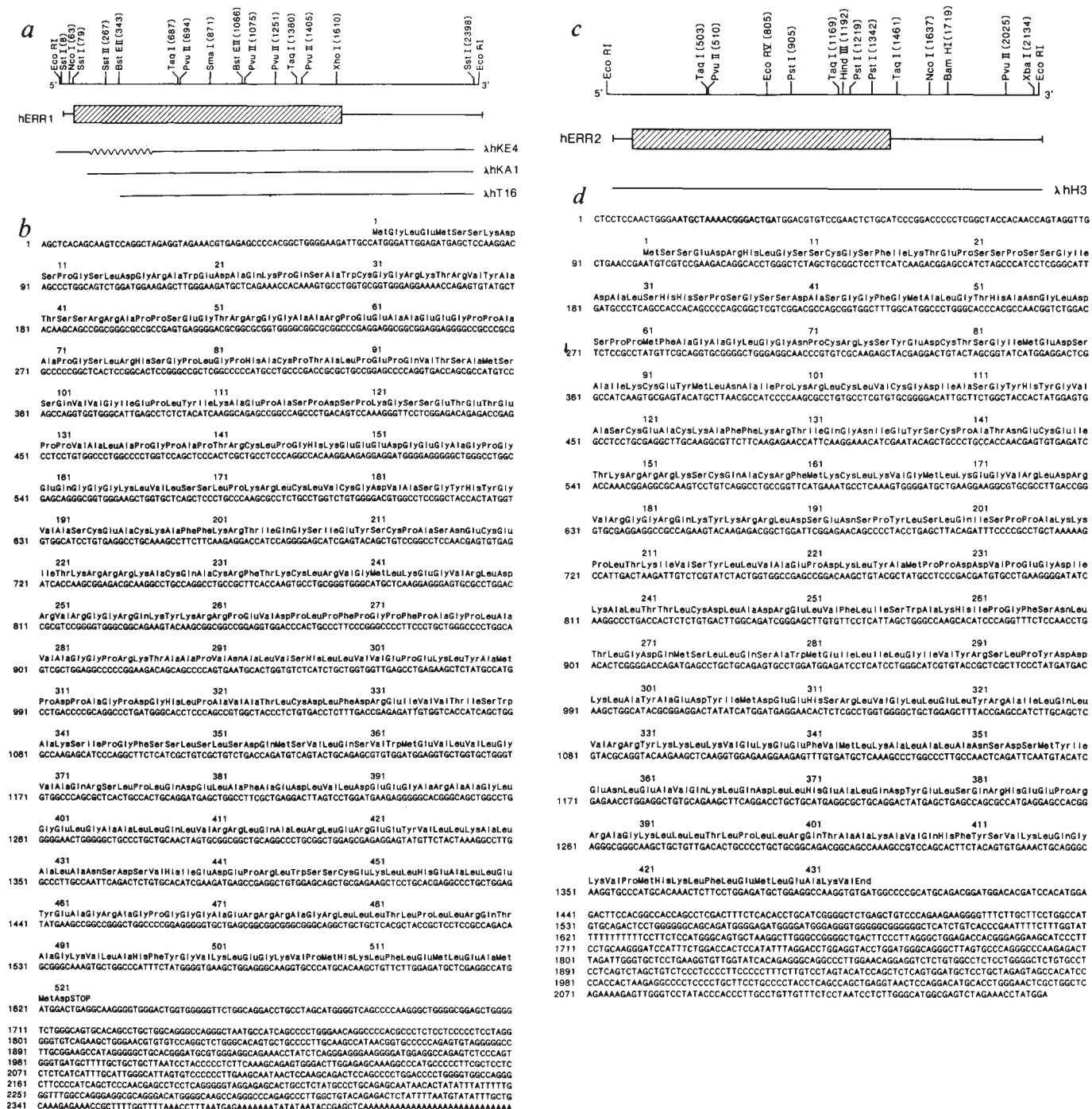


Fig. 1 Restriction map and DNA sequence and predicted amino-acid sequence of hERR1 and hERR2. **a**, The composite cDNA for hERR1 is represented at the top, with non-coding (thin line) and coding (stippled portion) sequences indicated. Common six-nucleotide restriction enzyme sites are drawn above the linear map. Overlapping cDNA inserts used to determine the sequence are shown. The wavy line near the 5' end of hKE4 indicates divergent sequence. **b**, Nucleotide sequence of the composite hERR1 cDNA with the deduced amino acids given above the long open reading frame. **c**, Schematic representation of hERR2 cDNA; some common restriction enzyme sites are indicated. The stippled box represents the predicted open reading frame. **d**, The complete nucleotide sequence of hERR2 is shown with the predicted amino-acid sequence given above the long open reading frame. A short open reading frame in the 5' untranslated region is shown in bold type.

Methods. The clone hT16 was isolated from a human testis λ gt10 cDNA library (Clontech) using a nick-translated¹⁸ 446-base-pairs (bp) *Bgl*I-*Bam*HI fragment isolated from pER945, a linker-scanning mutant¹² derived from the hER cDNA (V.G. and R.M.E., unpublished data). The hybridization mixture contained 35% formamide, 1× Denhardt's 5×SSPE (150 mM NaCl, 10 mM NaH₂PO₄, 1 mM EDTA), 0.1% sodium dodecyl sulphate (SDS), 100 μ g ml⁻¹ denatured salmon sperm DNA and 10⁶ c.p.m. ml⁻¹ of ³²P-labelled *Bgl*I-*Bam*HI fragment (>10⁸ c.p.m. μ g⁻¹). Duplicate nitrocellulose filters were hybridized at 42°C for 16 h, washed three times for 20 min each in 2×SSC, 0.1% SDS (1×SSC: 150 mM NaCl, 15 mM sodium citrate) at 55°C and autoradiographed at -70°C with an intensifying screen. The clones hKE4 and hKA1 were isolated from a human kidney λ gt10 cDNA library¹⁹ using the nick-translated insert from hT16. The clone hH3 was isolated from a human heart λ gt11 cDNA library using a nick-translated 700-bp *Eco*RI-*Sma*I fragment representing the 5' portion of hKA1. For this screening, the hybridization mixture was modified to 50% formamide and washing conditions to 2×SSC with 0.1% SDS at 68°C. The cDNA clones were digested with several restriction enzymes and the resulting fragments were subcloned in both orientations into the M13 sequencing vectors mp18 and mp19 and sequenced by the dideoxy procedure²⁰, and gaps or ambiguities resolved by the chemical cleavage method²¹ DNA sequences were compiled and analysed by the programs of Devereux *et al.*²² and Staden²³.

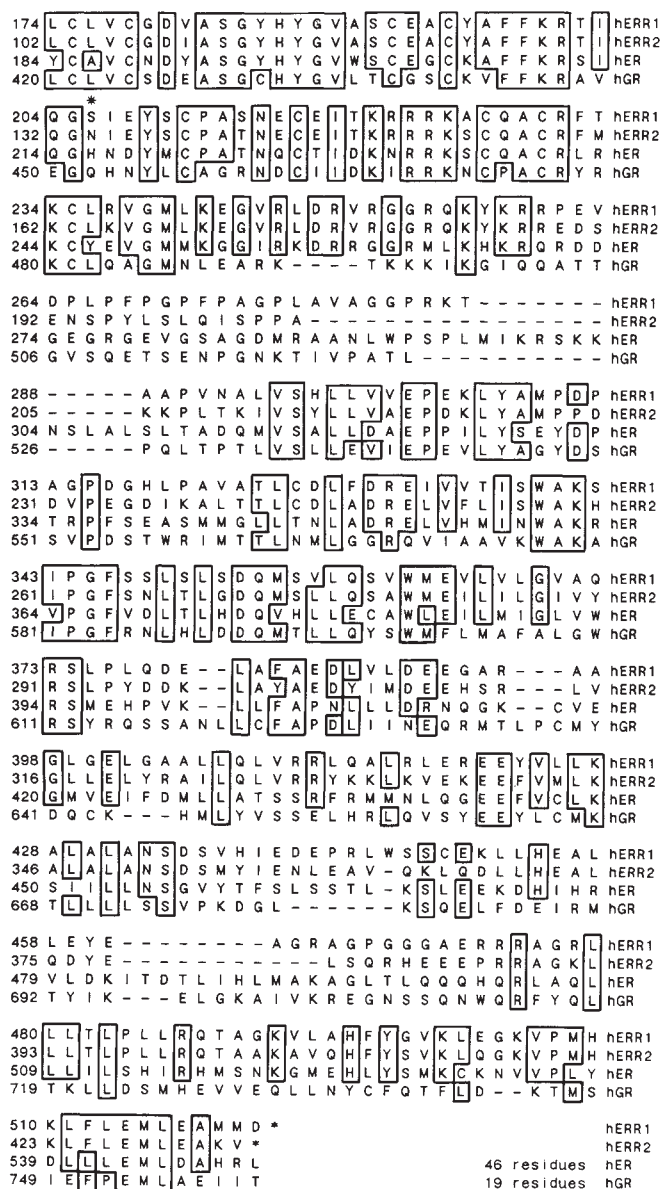


Fig. 2 Amino-acid sequence comparison between the carboxy-terminal regions of hERR1, hERR2, the human oestrogen and glucocorticoid receptors. The four amino-acid sequences were aligned for maximum similarity by introducing gaps as indicated by hyphens. Number were taken from Fig. 1 for hERR1 and hERR2, from Hollenberg *et al.*¹ for hGR and Green *et al.*² for hER. Amino-acid residues matched in at least three of the polypeptides are boxed. The asterisk above residue 206 of hERR1 indicates the position of the histidine residue present in the hER sequence but absent in both hERR1 and hERR2 sequences.

because this is the first ATG triplet that appears downstream from an in-frame terminator TGA (nucleotides 31–33). An open-reading frame containing 433 amino acids encodes a polypeptide of M_r 47.6K and is followed by a 3'-untranslated region of 752 nucleotides.

The predicted hERR1 and hERR2 polypeptides contain the characteristic domain features of steroid receptors. Amino-acid sequence comparison of hERR1 and hERR2 shows different amino termini with no similarity with other classes of receptor (data not shown), which agrees with the hypervariable sequence already demonstrated in this region^{5,8,10}. Sequence alignment of hERR1, hERR2, hER and hGR (Fig. 2) gives the greatest similarity in the 66 amino acids of the DNA-binding domain^{12,14}

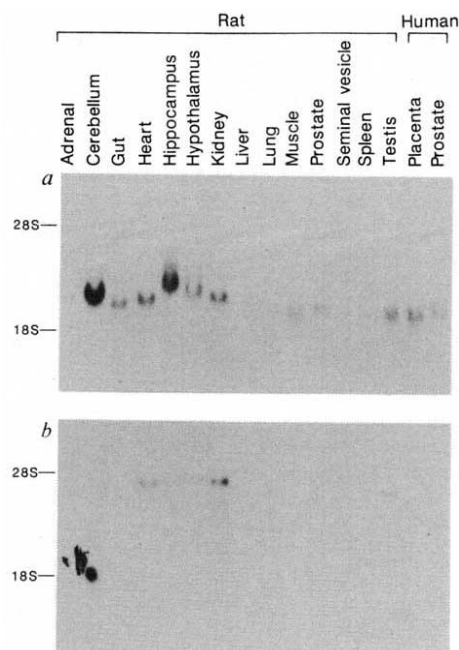


Fig. 3 Northern blot hybridization analysis of *a*, hERR1 and *b*, hERR2 mRNAs in rat and human tissues.

Methods. Total RNA was isolated using guanidine thiocyanate²⁴, separated on 1% agarose-formaldehyde gel, transferred to nitrocellulose, and hybridized under stringent conditions using a nick-translated *EcoRI*-*SmaI* fragment from λ hKA1 (*a*) and a nick-translated 1,192-bp *EcoRI*-*HindIII* fragment from λ hH3 (*b*). Lane loading of total RNA was 20 μ g. Migration of ribosomal RNAs (28S and 18S) are indicated for size markers. The nitrocellulose filters were autoradiographed at -70°C with an intensifying screen for 24 h (*a*) and 6 d (*b*). Apparent difference in migration rate of the mRNA in (*a*) is a gel artefact.

of the steroid hormone receptors. Between amino acids 175 and 240 of hERR1 there is 91% amino-acid identity with hERR2, 68% with hER, and 56% with hGR. The positions of the nine cysteine residues are strictly conserved but the absence of a histidine residue at position 206 of hERR1, and at position 134 of hERR2, marks a major difference with previously described steroid hormone receptors. It was originally thought that this histidine residue could be involved with the conserved cysteine residues in the formation of a DNA-binding finger. The recent demonstration that this histidine residue is also absent in the corresponding amino-acid sequence of the vitamin D receptor¹⁷, another member of the ligand-binding transactivation factor superfamily, suggests that Zn^{2+} atoms might interact exclusively with cysteine residues to coordinate the formation of the proposed DNA-binding fingers present in those proteins. The putative steroid-binding domain, positioned between amino acids 295 and 521 of hERR1, shows 63% identity when compared to hERR2, 36% to hER and 28% to hGR.

Steroid receptors are expressed in characteristic tissue-specific patterns that directly correlate with their primary physiological effects. To study the distribution of these putative receptors, total RNA was isolated from a variety of rat and human tissues and fractionated by formaldehyde-agarose gel electrophoresis and transferred to nitrocellulose filters. Using λ hKA1 as a probe, a 2.6-kb mRNA encoding hERR1 was detected in all rat and human tissues analysed, in surprisingly large amounts in the cerebellum and hippocampus, with the lowest concentrations being present in the liver, lungs, seminal vesicles and spleen (Fig. 3*a*). Thus, the hERR1 gene is widely and abundantly expressed, particularly in the rat central nervous system. Unlike the hERR1 mRNA expression pattern, the distribution of the mRNA encoding the hERR2 protein is restricted to very low

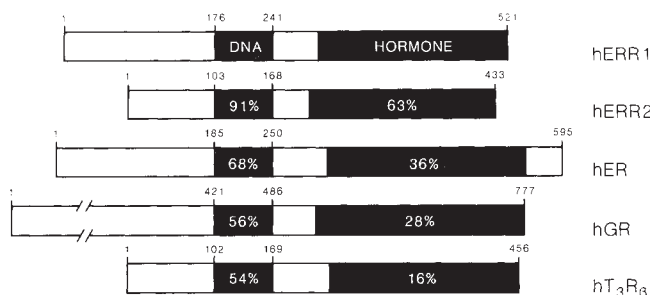


Fig. 4 Schematic amino-acid comparisons between hERR1 and hERR2, hER, hGR and human thyroid hormone receptor (hT₃R_β). Amino-acid sequences have been aligned schematically according to the functional domain structure of the steroid and thyroid hormone receptors superfamily. The percentage of amino-acid identity of each receptor with hERR1 is indicated inside each domain. The amino-acid position of each domain boundary is shown for each receptor.

concentrations in a few specific tissues (Fig. 3b). Using a probe derived from clone λhH3, a 4.8-kb mRNA was found in kidney, heart, testis, hypothalamus, hippocampus, cerebellum and rat prostate, with no detectable hybridization in human placenta or prostate. Considering the difference in exposure time and the resulting signal intensity, concentration of hERR2 mRNA are ~10–100-fold lower hERR1 mRNA.

Earlier studies have indicated that the degree of similarity of the ligand-binding domains between steroid hormone receptors reflects a structural similarity in their ligands. For example, the hGR, hMR and hPR have 56% identity in their ligand-binding domains⁵, and bind closely related hormones. Indeed, the hMR binds glucocorticoids with the same affinity as aldosterone, and also progesterone with relatively high affinity⁵. For the hERR gene products, amino-acid sequences are relatively more similar to hER, being 70% similar in the DNA-binding region and 36% similar in the steroid-binding region (Fig. 4). This similarity is lower than observed between hGR, hMR and hPR, and would thus predict that the putative hERR proteins interact with a class of steroid hormones distinct from the oestrogens. Preliminary steroid binding studies, using the products of *in vitro* translation of capped SP6 RNA produced from hERR1 and hERR2 coding sequences or from expression of the two cDNAs in COS-1 cells^{5,12}, have failed to demonstrate binding with any major classes of steroids, including oestrogens and androgens.

The tissue distribution of hERR1 and hERR2 mRNAs suggests that each putative receptor exerts distinct biological control. The functions of these steroid hormone receptors could have been overlooked if their activities had erroneously been attributed to other receptors, with differences classified as atypical effects. Thus, the isolation of novel steroid hormone receptor cDNAs is a step towards identifying a new hormone response system.

We thank our colleagues of the Evans laboratory for helpful discussions, Drs G. L. Greene for the gift of the human oestrogen receptor cDNA, G. I. Bell and L. A. Leinwand for providing human kidney and heart cDNA libraries, respectively. V.G. is a Centennial fellow of the Medical Research Council of Canada. N.Y. is a predoctoral trainee supported by a training grant to the Department of Chemistry, University of California, San Diego; R.M.E. is an Investigator of the Howard Hughes Medical Institute. This work was supported by the Howard Hughes Medical Institute and by grants from the National Institute of Health. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries under the accession number Y00290.

Received 29 September; accepted 28 October 1987.

1. Hollenberg, S. M. *et al. Nature* **318**, 635–641 (1985).
2. Green, S. *et al. Nature* **320**, 134–139 (1986).
3. Greene, G. L. *et al. Science* **231**, 1150–1154 (1986).

4. Misrachim, M. *et al. Biochem. biophys. Res. Commun* **143**, 740–748 (1987).
5. Arriza, J. L. *et al. Science* **237**, 268–275 (1987).
6. Miesfeld, R. *et al. Cell* **46**, 389–399 (1986).
7. Danielsen, M., Northrop, J. P. & Ringold, G. M. *EMBO J.* **5**, 2513–2522 (1986).
8. Krust, A. *et al. EMBO J.* **5**, 891–897 (1986).
9. Maxwell, B. L. *et al. Molec. Endocrinol.* **1**, 25–35 (1987).
10. Loosfelt, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 9045–9049 (1986).
11. Weiler, I. J., Lew, D. & Shapiro, D. J. *Molec. Endocrinol.* **1**, 355–362 (1987).
12. Giguere, V., Hollenberg, S. M., Rosenfeld, G. M. & Evans R. M. *Cell* **46**, 645–652 (1986).
13. Kumar, V., Green, S., Staub, A. & Chambon, P. *EMBO J.* **5**, 2231–2236 (1986).
14. Hollenberg, S. M., Giguere, V., Segui, P. & Evans, R. M. *Cell* **49**, 39–46 (1987).
15. Miesfeld, R., Godowski, P. J., Maler, B. A. & Yamamoto K. R. *Science* **236**, 423–427 (1987).
16. Kozak, M. *Nature* **308**, 241–246 (1984).
17. McDonnell, D. P. *et al. Science* **235**, 1214–1217 (1987).
18. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. J. *J. molec. Biol.* **113**, 237–251 (1977).
19. Bell, G. I. *et al. Nucleic Acids Res.* **14**, 8427–8446 (1986).
20. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
21. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
22. Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1985).
23. Staden, R. *Nucleic Acids Res.* **10**, 2951–2961 (1982).
24. Chirgwin, J. M., Przybyla, A. F., McDonald, R. J. & Rutter, W. F. *Biochemistry* **18**, 5294–5299 (1980).

Structure and expression of the human θ_1 globin gene

Shih-Lan Hsu*, Jon Marks†§, Jeng-Pyng Shaw†§, Ming Tam*, Douglas R. Higgs‡, Chien Celia Shen† & Che-Kun James Shen†||

* Institute of Molecular Biology, Academia Sinica, Taiwan, Republic of China

† Department of Genetics, University of California, Davis, California 95616, USA

‡ MRC Molecular Haematology Unit, Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

The recently identified θ -globin gene subfamily consists of the θ_1 -globin gene located downstream from the α_1 -globin gene^{1–5}, and several other members including at least one truncated, processed pseudogene $\psi\theta_2$ (refs 1, 6). Unlike the θ_1 -globin genes of the rabbit³ and galago⁴, the structure of these genes in the orangutan and baboon^{1,5} and their flanking regions show no apparent defects that would prevent their expression. Both θ_1 -globin genes are split into three exons with the potential to code for a polypeptide of length 141 amino acids. Besides differing by 26% in replacement-site substitutions, the θ_1 and α_1 -globin genes of the orangutan and baboon also differ in their promoter structures, in the use of TGA versus TAA as the termination codon, and in the use of AGTAA versus AATAAA as the polyadenylation signal^{1,5}. In contrast, the two θ_1 -globin genes from primates only differ by 1.7% in the replacement-site substitutions⁵. Here we present the complete DNA sequence of a cloned θ_1 -globin gene of humans, and show that it contains no apparent defects that would abolish its expression. Furthermore, by primer extension of single-stranded oligonucleotide probes, we show that the θ_1 -globin gene of humans is transcribed in an erythroleukemia cell line K562. Three messenger RNA species were detected, with 5'-ends mapping to ~70 base pairs (bp) downstream from a TATA promoter sequence, at 8 bp downstream from a GGGCGG promoter sequence and at 40 bp upstream from the ATG initiation codon, respectively. Haemin treatment of the K562 cells slightly enhances the level of the longest θ_1 -transcript. Our results provide strong evidence that the θ_1 -globin gene of humans is transcriptionally active in cells of erythroid origin, and suggests the presence of a functional θ_1 -polypeptide in specific cells, possibly those of early erythroid tissue.

§ Present addresses: Department of Anthropology, Yale University, New Haven, Connecticut 06520, USA (J.M.). Department of Pathology, Stanford School of Medicine, Stanford, California 94305, USA (J.-P. S.).

|| To whom correspondence should be addressed.


```

      |      |      |      |      |      |      |      |      |      |      |
CCACTGCACCTACCGCACCGGCGCAATTTTGTGTTTTAGTAGAGACTAAATACATATAGTGAACACCTAAGACGGGG (80)
      |      |      |      |      |      |      |      |      |      |      |
GGCCTTGGATCCAGGCGGATTCAGAGGCGCCGGCTCGGAGCTGTCCGAGATTGAGCGCGCGCGGTCCGGGATCTCCGAC (160)
      |      |      |      |      |      |      |      |      |      |      |
GAGGCGCTGACCCCGCGCGGCGAAGCTGCGGCGCGCGCCCTGAGGCGCGCGGACCCCTGCGCGGTCCGCGCAGG (240)
      |      |      |      |      |      |      |      |      |      |      |
CGCAGCGGGGTTCGACGGGCGCGGGGTTCCAGCGGGGGTATGGCGCTGTCCGCGGAGGACCGCGCGGTGGTCCGCGCC (320)
      |      |      |      |      |      |      |      |      |      |      |
      |      |      |      |      |      |      |      |      |      |      |
TCTCGAAGAGCTGGGAGCAACCTCGGCTCTACACGACAGAGGCGCTGGAAAGTGTCCGCGCGGTGGCGCGCGCC (400)
      |      |      |      |      |      |      |      |      |      |      |
      |      |      |      |      |      |      |      |      |      |      |
euTrpLysLysLeuGlySerAsnValGlyValTyrThrThrGluAlaLeuGluAr
      |      |      |      |      |      |      |      |      |      |      |
CCCAGGGGCTCTCTCCCAAGCCCCCGGACGCGCTCACCACAGTTCCTTCGCGAGGACCTTCTGGCTTTCCCGC (480)
      |      |      |      |      |      |      |      |      |      |      |
      |      |      |      |      |      |      |      |      |      |      |
gThrPheLeuAlaPheProAl
      |      |      |      |      |      |      |      |      |      |      |
CACGAAGACCTACTTCTCCACCTGGAACCTGAGCGCGGGTCTCTCACAAGTCAGAGGCCACGGCCAGAGGTGGCGGACG (560)
      |      |      |      |      |      |      |      |      |      |      |
      |      |      |      |      |      |      |      |      |      |      |
aThrLysThrTyrPheSerHisLeuAspLeuSerProGlySerSerGlnValArgAlaHisGlyGlnLysValAlaAspA
      |      |      |      |      |      |      |      |      |      |      |
CGCTGACCTCCCGCTGGAAGCGCTGGACGACCTACCCACGCGCTTCGCGCGTGGAGCCACCTGCACCGGTGCCAGCTG (640)
      |      |      |      |      |      |      |      |      |      |      |
      |      |      |      |      |      |      |      |      |      |      |
laLeuSerLeuAlaValGluArgLeuAspAspLeuProHisAlaLeuSerAlaLeuSerHisLeuHisAlaCysGlnLeu
      |      |      |      |      |      |      |      |      |      |      |
CGAGTGGACCGCGGCTCCAGGTGAGCGGCTGCGCTGCTGCGCGCTGTCTCCCGGAGGCGCGCGCGGTGGG (720)
      |      |      |      |      |      |      |      |      |      |      |
      |      |      |      |      |      |      |      |      |      |      |
ArgValAspProAlaSerPheGln
      |      |      |      |      |      |      |      |      |      |      |
CGGGGGGCTGCGGGGCGGTGCGAGCGAGTGAGCTTGAGCGCTCGCGAGCTCTGGGCGACTGCTGCTGTAACC (800)
      |      |      |      |      |      |      |      |      |      |      |
      |      |      |      |      |      |      |      |      |      |      |
LeuLeuGlyHisCysLeuLeuValThr
      |      |      |      |      |      |      |      |      |      |      |
CTCGCCCGGCACTACCCGAGACTTCAGCGCGCGCTGAGCGCTGCTGGCAAGTTCCTGAGCCAGCTTATCTCGGC (880)
      |      |      |      |      |      |      |      |      |      |      |
      |      |      |      |      |      |      |      |      |      |      |
LeuAlaArgHisTyrProGlyAspPheSerProAlaLeuGlnAlaSerLeuAspLysPheLeuSerHisValIleSerAl
      |      |      |      |      |      |      |      |      |      |      |
GCTGGTTCCGAGTACCGCTGAACCTGCGGTGGTGGCGCGGATCCCGAGGCGACCTCCCGGTGTTTGAATAAGCC (960)
      |      |      |      |      |      |      |      |      |      |      |
      |      |      |      |      |      |      |      |      |      |      |
aLeuValSerGluTyrArgTer
      |      |      |      |      |      |      |      |      |      |      |
TCTCCAGGAGCAGCCTCTCTCCGCTGCTCTCTCGAGTTCAGGACCGGAGGAGGCGG (1020)

```

Fig. 1 Nucleotide sequences of the θ_1 -globin gene of humans. Sequences begin 280 bp upstream of the ATG codon, and extend for 48 bp past the putative polyadenylation site (base 972). The promoter-sequence CCAAT (bases 23–27), TATA (bases 57–61) and GGGCGG (bases 177–182) are boxed. The polyadenylation signal AGTAAA and the putative polyadenylation site are underlined. The amino-acid translation is given below the nucleotide sequence.

Methods. The $\alpha 1$ - and $\theta 1$ -globin genes of humans are contained within a single 8 kilobase *Bgl*II fragment^{1,8}. This fragment has been cloned in the cosmid *C* $\alpha 3'$ Bg (ref. 9). The 900 bp *Bam*HI fragment of *C* $\alpha 3'$ Bg and the DNA regions immediately flanking it were sequenced by the method of Maxam and Gilbert¹⁰.

The nucleotide sequences of the θ_1 -globin gene of humans is shown in Fig. 1. The sequence elements characteristic of the θ_1 -globin genes of primates^{1,2,5} are all found within or flanking the θ_1 -globin gene of humans (Fig. 1). These include the CCAAT and TATA boxes that are displaced further upstream from the ATG initiation codon, the TGA termination codon and the AGTAAA polyadenylation signal. The unique sequence organization of the DNA region near the CCAAT and TATA boxes of the θ_1 -globin genes of orang-utan and baboon, which is not found in the α -globin genes, is well conserved in the θ_1 -globin gene of humans. As predicted from studies of its orthologous genes in orang-utan and baboon^{1,5}, the θ_1 -globin gene of humans contains no apparent gene defects such as premature termination codons, or other mutations that would prevent its expression. The comparison of replacement-site substitutions and silent-site substitutions among the human α - and ζ -globin genes and the three primate θ_1 -globin genes (Table 1) suggest that the θ_1 -globin genes started to diverge from the α -globin genes 260–280 million years ago.

Our previous studies revealed little transcription of the θ_1 -globin gene of the orang-utan transfected into the HELA cells or monkey COS cells^{1,5}. We have now looked for the expression of the human θ_1 -globin gene in the erythroleukaemic cell line K562¹¹. All of the known globin genes of human, except for the β -globin gene of adults, are transcriptionally active and

produce globin polypeptides in K562 cells^{12–17}. The transcription of these globin genes in K562 cells are enhanced by the treatment with haemin^{12,13}.

Primer extension has been used to investigate the possible transcription of the θ_1 -globin gene of humans in K562 cells. We first synthesized a 36-nucleotide (nt) long, single-stranded oligonucleotide primer, 5'-CCGGTCTCTCCGCGACAGCG-CCATCCCCGCGCTGGA-3'. This probe is complementary to the sense strand of the θ_1 -globin gene of human between the positions 12 bases upstream and 24 bases downstream of the ATG codon (Figs 1 and 2a). This probe was chosen because its 3'-portion does not hybridize with other known globin-mRNAs of humans, and thus could not be extended by the reverse transcriptase. The oligonucleotide primer-extension reactions are indeed highly gene-specific. Both the α - and γ -oligonucleotide primers are extended by reverse transcriptase to the expected lengths after hybridization to K562 RNAs. The 3'-ends of the extension products of the α -probe (Fig. 2b) and the γ -probe (data not shown) coincide with the known capping sites of the α - and γ -globin mRNAs, respectively^{18,19}. Their mRNA levels are also enhanced by haemin treatment of the K562 cells (Fig. 2b), in agreement with previous reports^{12,13}. In the experiment shown in Fig. 2b, the α -mRNA level is increased about sevenfold after haemin induction (compare lane 2 to lane 1, Fig. 2b).

Interestingly, θ_1 -transcripts are also present in the K562 cells. As seen in lane 3 of Fig. 2b, three DNA bands corresponding to lengths of about 177, 115 and 64 nt, respectively, appear after hybridization of the θ_1 -specific probe to K562 poly(A) RNA, and extension by the reverse transcriptase. The level of the θ_1 -mRNA in K562 cells is between one-tenth to one-fifth of the α -globin mRNAs. When compared to its effect on the transcription of α -globin genes, the effect of haemin treatment on the θ_1 -mRNA level is relatively small. Whereas the 177 nt primer-extension band is enhanced twofold with haemin induction, the bands corresponding to 115 and 64 nt are relatively unchanged or even decreased (compare lanes 3 and 4, Fig. 2b).

These results show that the promoter of the θ_1 -globin gene

Table 1 Percentage comparison of replacement- and silent-site substitutions of different primate α -globin-like genes

human ζ	38 (60)			
human θ_1	28 (67)	45 (72)		
orang-utan θ_1	26 (51)	42 (54)	0.39 (3.5)	
baboon θ_1	21 (55)	41 (68)	2.4 (23)	1.7 (22)
	human α	human ζ	human θ_1	orang-utan θ_1

The numbers represent the percentages of replacement-site substitutions and silent-site substitutions (in parentheses) among different pairs of θ_1 , α , and ζ -globin genes. The calculations were made according to Perler *et al.*²⁴.

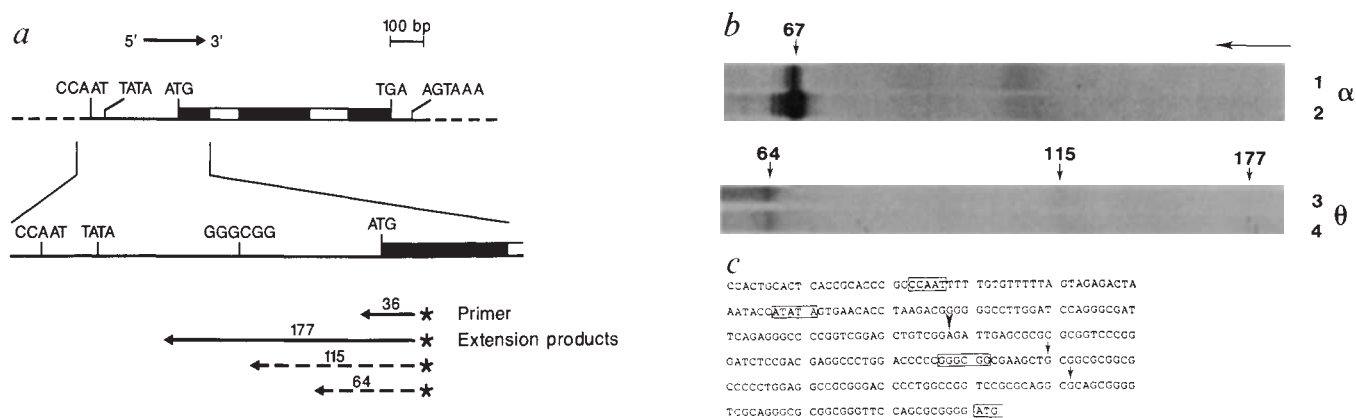


Fig. 2 Primer-extension analysis of the transcription of the θ_1 -globin gene of humans in K562 cells. **a**, Strategy. Shown at the top is the map of the θ_1 -globin gene of humans, with the three exons represented by filled boxes. Below the top map is the DNA region surrounding ATG codon with the CCAAT, TATA and GGGCGG promoter boxes indicated. The vertical arrows below the middle map indicate the putative transcriptional initiation sites as analysed by primer extension (see text and Fig. 2b). The θ_1 -specific oligonucleotide primer used to hybridize with the RNA samples, and its extension products are represented by horizontal arrows with their lengths (nt) indicated. The asterisk shows the labelled 5' ends. **b**, The mapping results locate three putative initiation sites of θ_1 -globin gene in K562 cells to 153 bp, 90 bp and 40 bp, respectively, upstream of the ATG codon. **c**, Gel electrophoresis of primer extension products. The direction of electrophoresis is from right to left. The numbers denote the lengths (nt) of the extension products, as determined by DNA sequence markers (not shown). These extension products were not seen when transfer RNA was substituted for the K562 RNA. Lane 1, α -probe with poly(A) RNA from uninduced K562 cells; lane 2, α -probe with poly(A) RNA from haemin-induced K562 cells; lane 3, θ_1 -probe with poly(A) RNA from uninduced K562 cells; lane 4, θ_1 -probe with poly(A) RNA from haemin-induced K562 cells. **c**, Transcriptional initiation map of the θ_1 -globin gene of human in K562 cells. The nucleotide sequence upstream from the ATG initiation codon is shown with the three promoter elements CCAAT, TATA and GGGCGG boxed. The three putative sites of transcriptional initiation of the θ_1 -globin gene in K562 cells are indicated by the vertical arrows. **Methods.** Three single-stranded oligonucleotide probes complementary to specific regions of the sense strands of the θ_1 , α - and γ -globin gene of human s, respectively, were synthesized by oligosynthesizer (Pharmacia) and purified. The probe for the θ_1 -globin gene is from base 259 to base 304 of Fig. 1. This would span the 5'-portion of θ_1 -coding region and the 12 bp upstream of it. The probes for the α -globin gene is 30 nt long, and spans from one nucleotide upstream of its ATG codon¹⁸ to 29 nucleotides downstream of it. The probe for the γ -globin gene is 34 nt long and spans from 14 nucleotides upstream of the ATG codon¹⁹ to 20 nucleotides downstream of it. The primer extension reactions were carried out as described previously^{20,21}. The probes were first labelled with ³²P at their 5'-ends by T4 polynucleotide kinase (Boehringer). After hybridization 5 ng of each primer with either 20 μ g total RNA or 2 μ g poly(A) RNA from K562 cells, or a haemin-induced K562 cells (0.05 mM haemin for 5 days), the primers were extended by reverse transcriptase from the avian myeloblastosis virus (AMV) (Boehringer), and the products were analysed by electrophoresis on denaturing 8% polyacrylamide, 7M urea gels and X-ray autoradiography.

functions in K562 cells and responds weakly to haemin induction. Assuming no splicing in the 5' untranslated region of θ_1 transcripts, the largest primer-extension product suggests the existence of one major transcriptional initiation site at ~70 bp downstream of the TATA promoted box (Fig. 2c). If the two smaller primer extension products were not generated by the degradation or processing of large RNAs, or as the result of RNA secondary-structure formation during the primer-extension reactions, they also suggest initiation sites at 90 bp and 40 bp upstream of the ATG initiation codon, respectively. The initiation of the longest transcript may be directed by the CCAAT and TATA boxes, although the distance (70 bp) between the TATA box and the initiation site is longer than in most of the other eukaryotic RNA polymerase II-dependent genes. The GGGCGG promoter box²² at ~100 bp upstream the ATG codon may be responsible for the initiation of one of the two minor transcripts, whose 5'-end maps at 8 bp downstream of the GC box (Fig. 2c). The third putative site of initiation, which is located within a GC-rich region, has previously been detected with the S1 nuclease mapping technique²³.

Our demonstration that the θ_1 -globin gene of humans is functional in K562 cells is consistent with a previous study which showed that this gene is transcribed in the fetal liver and fetal yolk sac of humans, but not in several adult erythroid or non-erythroid tissues²³. Although this previous study failed to detect the two more upstream transcription initiation sites mapped by us (see Fig. 2), due to the limitations of the 5'-S1 probe, their 3'-S1 mapping did show clearly that the θ_1 transcript is processed and polyadenylated efficiently *in vivo*²³, at a site predicted previously by the sequence analysis^{1,5}. The demonstration of θ_1 -mRNA synthesis in cells from humans significantly strengthens the case for the possible existence of a θ_1 -polypeptide *in vivo*, whose function is not yet known. There are no

apparent abnormal developmental patterns associated with individuals heterozygous²⁵ or homozygous²⁶ for θ_1 deletion.

This research has been supported by the National Council of Science, Republic of China, the MRC, and the NIH. C.-K. J.S. is supported by a NIH Research Career Development Award and a Fogarty Senior International Fellowship. We thank Nick Proudfoot and his colleagues, Dr Bon-Chu Chung for help and discussions L.-F. Liu for technical assistance, Drs Bon Chu Chung, Tao-Shih Hsieh, J.-L. Huang, Simon Hui, T.-T. Kuo, Gloria Li, Leroy Liu, Robert Lo, Ming Tam, David Tu, and Jim Wang.

Received 9 August; accepted 30 October 1987.

- Marks, J., Shaw, J.-F. & Shen, C.-K. *J. Nature* **321**, 785-788 (1986).
- Marks, J. *et al. Cold Spring Harb. Symp. quant. Biol.* **51**, 499-508 (1986).
- Cheng, J.-F., Raid, L. & Hardison, R. C. *J. biol. Chem.* **261**, 839-848 (1986).
- Sawada, F. & Schmid, C. W. *J. molec. Biol.* **192**, 693-709 (1986).
- Shaw, J.-P., Marks, J. & Shen, C.-K. *J. Nature* **326**, 717-720 (1987).
- Shaw, J.-P., Marks, J., Mohandas, T., Sparkes, R. & Shen, C.-K. *J. in Haemoglobin Switching V* (eds Stamatoyannopoulos, G. & Nienhuis, A. W.) (in the press).
- North, G. *Nature* **326**, 641 (1987).
- Lauer, J., Shen, C.-K. J. & Maniatis, T. *Cell* **20**, 119-130 (1980).
- Nicholls, R. D., Fischel-Ghodsian, N. & Higgs, D. R. *Cell* **49**, 369-378 (1987).
- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
- Lozzio, B. & Lozzio, C. *Leuk. Res.* **3**, 363-370 (1979).
- Dean, A., Ley, T., Humphries, R., Fordis, M. & Schechter, A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5515-5519 (1983).
- Charnay, P. & Maniatis, T. *Science* **220**, 1281-1283 (1983).
- Rutherford, T., Clegg, J. & Weatherall, D. *Nature* **280**, 164-165 (1979).
- Dean, A., Erard, F., Schneider, A. & Schechter, A. *Science* **212**, 459-461 (1981).
- Rutherford, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 348-352 (1981).
- Karlsson, S. & Nienhuis, A. W. *Rev. Biochem.* **54**, 1071-1108 (1985).
- Liebhafner, S. A., Goossens, M. J. & Kan, Y. W. *Nature* **290**, 26-29 (1981).
- Slightfoot, J. L., Blechl, A. E. & Smithies, O. *Cell* **21**, 627-638 (1980).
- Treisman, R., Proudfoot, N. J., Shander, M. & Maniatis, T. *Cell* **29**, 903-911 (1980).
- Hess, J. F. *et al. J. molec. Biol.* **184**, 7-21 (1985).
- Kadonaga, J. T., Jones, K. A. & Tijan, R. *Trends biochem. Sci.* **11**, 20-23 (1986).
- Leung, S.-O., Proudfoot, N. J. & Whitelaw, E. *Nature* **329**, 551-556 (1987).
- Perler, F. *et al. Cell* **20**, 553-566 (1980).
- Fei *et al. Br. J. Haemat.* (in the press).
- Fischel-Ghodsian N., Beyer, E. C. & Higgs, D. R. *Nature* (in the press).

Innovations for the new year

This week's issue looks at new laboratory products to make research safer, faster and easier, including a robotic arm, a see-through incubator and a pocket photometer.

BETHESDA Research Laboratories has a new electrophoresis system for **miniature gels** measuring 5.7×8.3 cm (*Reader Service No. 100*). The Horizon 58 unit features a slide-out gel-casting dam that eliminates the use of tape, and a thick lift-out plastic gel deck that ensures warp-free casting without precooling the agarose.



BRL's mini-gel unit stacks for convenience.

The buffer tray holds 100 ml of buffer, and can be removed easily for buffer disposal. The gel casting unit includes gel-casting deck, dams, buffer tray, base unit, 8- and 14-well combs, and power cords. The unit can be stacked on top of the model 200 power supply, sold separately, to save space on the lab bench.

The American Type Culture Collection has announced its series of **workshops for 1988** (*Reader Service No. 101*). Topics for January through June include: an update on hybridomas and monoclonal antibodies, recombinant DNA technology, freezing and quality control of cell cultures and hybridomas, freezing and freeze-drying of microorganisms, and gas chromatographic analysis of cellular fatty acids for bacterial identification. The laboratory-intensive workshops last from two to five days, and count for up to three CEU credits. In April, the ATCC is also sponsoring a biotechnology patent conference, including discussions of national and international patenting requirements, and the current status of patenting and depositing animal varieties.

An **environmental cabinet** for breeding laboratory animals is available from Equipements Scientifiques & Industriels (*Reader Service No. 102*). The model E110S has a self-contained air flow system that comes in either a positive or negative pressure configuration. The positive pressure cabinet maintains a germ-free atmosphere for animals who require protection against pathogens, while the negative pressure cabinet can be used to protect the laboratory environment from pathogens carried by animals inside the cabinet.

Both models have a prefilter that screens out 90 per cent of particles over $5 \mu\text{m}$, and an absolute filter to remove 99 per cent of $0.3\text{--}5.0 \mu\text{m}$ particles. Each cabinet holds about 16 medium-sized cages, and has four separate compartments and eight doors.

DR Lange has a **pocket photometer** for water analysis, the LASA Aqua (*Reader Service No. 103*). The unit has 14 program filters for the analysis of ammonium, cadmium, calcium, chloride, chlorine, chlorine dioxide, chromium, copper, cyanide, fluoride, formaldehyde, hexacyanoferrates, iron, magnesium, nickel, nitrate, nitrite, ozone, phenols, phosphate, potassium, silver, sulphate, zinc and water-hardness. The appropriate filters needed to generate the wavelength suitable for the analysis of a particular ion



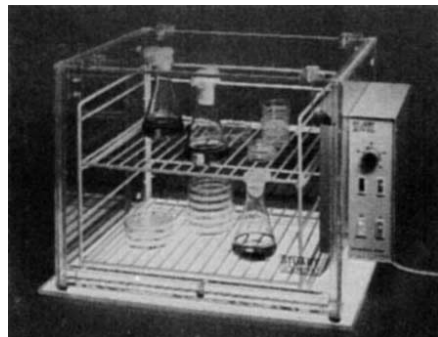
The DR Lange photometer fits in a pocket.

are put in place when one of the three operation buttons are pressed. The LASA Aqua is automatically calibrated, so no zero-adjustment using blank cuvettes is required.

A new brochure from Perkin-Elmer describes the MasterLab **robot system** for sample preparation and analysis (*Reader Service No. 104*). The free brochure outlines the capabilities of the MasterLab core system, which consists of a hinged-arm robot designed to emulate the human arm, a system controller, and Perkin-Elmer robot language. The MasterLab robot is designed for use in a chemical or pharmaceutical laboratory. It features a linear transport belt to allow the robot to move, and interchangeable tactile-sensing 'fingers' for grasping objects ranging from test tubes to microtitre plates. Liquid chromatographs, Mettler titrators,

UV-visible spectrophotometers and other instruments may be interfaced to the MasterLab system, and the robot can be programmed to perform sample injection. The robot can also perform weighing steps, pipetting, centrifugation, and mixing, using a vortex mixer or rotating sample racks.

For keeping an eye on samples under incubation, Stuart Scientific sells an **acrylic 60-litre incubator**, the model SI60



See what you grow with Stuart Scientific's incubator.

(*Reader Service No. 105*). The incubator opens at the top, front and base, and the wire rack may be removed for incubating entire instruments. For working with anaerobes, the unit is available with glove ports and inlets for introducing cables or gases. With the addition of a salt-tray humidifier, the incubator becomes a non-injection humidity chamber. Stuart Scientific says fan convection within the unit ensures uniform humidity and temperature fluctuation of less than 0.1°C .

ICN Biomedicals has a brochure devoted to **animal research diets** (*Reader Service No. 106*). The free brochure lists information on the 100 standard diets and the over 10,000 custom formulations sold by the company. The diets come in three pellet sizes, and are colour-coded to ease control of administration.

Laboratory personnel can be shielded from beta radiation from liquid radioactive waste with Rad-Can **beta shields** from Research Products International (*Reader Service No. 107*). The acrylic boxes come in two sizes: the \$75 (US) Rad-Can II holds a one gallon bottle, and the \$79 (US) Rad-Can III accepts a two-gallon carboy. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

Lack of mobility becomes critical

Richard Pearson

The north-south divide in the United Kingdom is a reality — at least where housing costs are concerned — and employers must cope with the results.

This year 250,000 people will move to a new home at the request of their employer. With a garage in some areas of London now costing £50,000 or more, which is more than a four-bedroom detached house in many parts of the country, it is not surprising that employers in many parts of the south-east find it difficult to move or to attract staff from less prosperous regions. But house price differentials are not the only cause of low mobility in the United Kingdom, disruption to family life, children's education and spouses' careers can be just as important, while the intrinsic merit of the new location is also critical.

For many employers, housing is the only apparent barrier and some continue to throw money at the growing problem of employee immobility. For those changing employer, assistance can, however, range from nothing — some employers do not even reimburse interview costs — to payment of legal and removal expenses, which is common in the public sector, as well as estate agents' fees plus, in some cases, payment of a disturbance allowance, costs of temporary accommodation and bridging finance.

The big money, however, is spent when the employer wants an individual to relocate within the company. In these cases the package can include all the above plus time off for house hunting, travel and accommodation for dependants, relocation expenses of up to 10 per cent or more of salary, a separate disturbance allowance and help with housing costs including subsidized mortgages, bridging finance, and "high housing cost" compensation whereby the employer pays the additional mortgage costs of an equivalent but higher priced house. There are now numerous relocation consultancies that will help lubricate this process for a fixed fee payable by the employer. They do this by offering to buy the house of the

employee about to move, assisting with briefings and house search in the new location and arranging mortgages. All of this is aimed at removing the stress of a move as well as providing a marginal cash incentive in some cases. The total costs of moving a senior manager can now be £15,000 or more. But not everyone pays on this scale and the solution to immobility is rarely a simple one of financing relocation.

A recent survey of 8,000 managers and professionals has shown the wide range of factors restraining mobility (Table 1). The respondents came from five major public and private sector employers, and nearly half of them had relocated at least once in the past ten years. Nevertheless, relocation remains rare, only 16 per cent having moved more than once in the last 10 years. Those in their twenties were more likely to have moved than their older counterparts and men were more likely to have moved than women. For two thirds, the attraction was career advancement, for the others it was salary increases. One in three had turned down a move in the past 10 years. The main reasons for their refusals were the maintenance of social and family relationships, affecting 2 out of 3, followed by a dislike of the new location. House prices came third and other important factors included schooling, spouses' jobs and careers, and salary and relocation costs (Table 1). Surprisingly, although spouses' careers were important, three quarters of those with working spouses, who may be male or female, maintained that their career took precedence over their spouse's, dual careers having yet to have a major impact. Although employees are becoming more selective about moving, they will nevertheless move if the conditions are right, with 40 per cent saying they expected to move in the next ten years and 60 per cent saying they would do so to secure promotion and a higher salary.

So where would they go? First of all, long-distance moves are not surprisingly far less popular than short-distance ones. There were also extreme aversions to moving to 'cities', 'urban areas', 'inner cities' and 'industrial areas'. In terms of specific locations, London was the least attractive location followed by Scotland and Northern Ireland, with other locations attracting far fewer negative attitudes (Table 2). The so-called north-south divide was, as expected, a barrier but not one of absolute importance, with many moves being within the north or within the

south. While moves from north to south were less popular than from south to north, the real barrier was into London.

While a number of big employers are relocating their offices out of London and the south-east to avoid recruitment difficulties, if they are seeking to move individual employees they usually have to do so within a fairly constrained set of locations. So what can they do to improve the mobility of their workforce? First, they can recognize different employees' susceptibility to moving. For example, the young are far more likely to be mobile, with willingness peaking in the mid-20s and declining thereafter.

But before any action can take place, someone has to admit to the problem on

Table 2 Least attractive areas (in rank order)

- 1 London
- 2 Scotland
- 3 Northern Ireland
- 4 North-west
- 5 Wales
- 6 North
- 7 South-east

behalf of the company. Too often the responsibility for mobility is devolved to the line managers involved who have to pick up the problem on an *ad hoc* and largely reactive basis. The first action is to establish clear managerial responsibility for mobility and then to understand why and where mobility usually takes place as well as whether it is actually needed or is simply an organizational habit or apparent convenience. Demand for mobility can often be reduced by the introduction of regional or localized career structures or ensuring that expansion takes place outside difficult or high-cost areas, where relocation may even be cost effective. A better understanding of staff needs and concerns will also help, as will ensuring that employees know what help is available to them when they are thinking about the move. Incentives for the mobile have to be reviewed, not only in terms of both immediate pay and relocation benefits, but also in terms of longer-term earnings and career payback. At the end of the day though, however good the relocation package, that in itself will not be an incentive to move, it can only be a compensation. Longer-term career and pay prospects will be the only real incentives. □

Table 1 Main reasons for deciding against a move

	per cent
Relatives/friends	64
Family against	59
Poor location	56
Housing cost	51
Salary too low	47
Dependent relatives	43
Schooling	41

These data and other figures in this article are from the IMS report, "Relocating Managers and Professional Staff" (1987).

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.